

Mean-field type modeling of nonlocal congestion in pedestrian crowd dynamics

Alexander Aurell ^{*} Boualem Djehiche [†]

February 1, 2017

Abstract: We extend the class of pedestrian crowd models introduced by Lachapelle and Wolfram (2011) to allow for nonlocal congestion and arbitrarily but finitely many interacting crowds. The new congestion feature grants pedestrians a ‘personal space’ where crowding is undesirable. We treat the model as a mean-field type game which we derive from a particle picture. Solutions to the mean-field type game are characterized via a Pontryagin-type Maximum Principle. The behavior of a crowd acting under nonlocal congestion is illustrated by a numerical simulation.

MSC 2010: 49N90, 60G09, 60H10, 60H30, 60K35

Keywords: Crowd dynamics, congestion, mean-field approximation, interacting populations, optimal control, mean-field type games

1. Introduction

When moving in a crowd, a pedestrian chooses its path based not only on its desired final destination but it also takes the movement of other surrounding pedestrians into account. The bullet points below are stated in [14] as typical traits of pedestrian behavior.

- Will to reach specific targets. Pedestrians experience a strong interaction with the environment.
- Repulsion from other individuals. Pedestrians may agree to deviate from their preferred path, looking for free surrounding room.
- Deterministic if the crowd is sparse, partially random if the crowd is dense.

These properties appear in classical particle models. Other authors advocate smart particle models that follow decision-based dynamics. In [14] some fundamental differences between classical and smart particle models are outlined. We list a few of them in Table 1.

^{*}Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden.
E-mail address: aaurell@kth.se

[†]Department of Mathematics, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden.
E-mail address: boualem@kth.se

Classical	Smart
Robust - interaction only through collisions	Fragile - avoidance of collisions and obstacles
Blindness - dynamics ruled by inertia	Vision - dynamics ruled at least partially by decision
Local - interaction is pointwise	Nonlocal - interaction at a distance

TABLE 1

A smart particle model lets pedestrians decide where to walk and with what speed. The choice is based on some rule that takes the available information into account such as the positioning and movement of other pedestrians. Although more realistic, this approach has complications. If pedestrian i moves, all pedestrians accessing information on i 's state might have to adapt their movements. The large number of connections where information is exchanged within a crowd is a computational difficulty.

The mean-field approach to modeling congestion in pedestrian crowds was introduced in [12], where pedestrians are treated as particles, following decision-based dynamics, trying to optimize their path by avoiding local congestion.

In a crowd of N pedestrians, an N -crowd from now on, pedestrian i with position $X^{i,N}$ controls its velocity such that its congestion risk measure, $J^{i,N}$, is minimized over a finite time horizon $[0, T]$. The risk measure penalizes congestion, energy waste and failure to reach a target area. In this paper we advocate for the use of the following nonlocal congestion term.

$$\mathbb{E}_N \left[\int_0^T \frac{1}{N-1} \sum_{j=1, j \neq i}^N \phi_r \left(X_t^{i,N} - X_t^{j,N} \right) dt \right], \quad (1.1)$$

where the function ϕ_r models the ‘personal space’ of a typical pedestrian and $X_t^{i,N} - X_t^{j,N}$ is the distance between pedestrian i and pedestrian j at time t . The personal space, ϕ_r , has support within a ball of radius r and hence a pedestrian is not effected by crowding outside it. For positive r , (1.1) is a weighted average of the crowding within the personal space. Connecting to the terminology in Table 1, we will refer to this as nonlocal congestion. Taking the limit $r \rightarrow 0$ corresponds to shrinking the personal space to a singleton and only pointwise crowding, that is collisions, will effect the pedestrian. This is called local congestion.

In emergency situations it is often in the interest of all pedestrians to get to a certain place, such an exit. In evacuation planning or crowd management at mass gatherings, it is in the interest of the planner to control the crowd along paths and towards certain areas. Common to such situations is the conflict between attractive locations and repulsive congestion in the crowd. Pedestrians acting under nonlocal congestion will order themselves more densely in such places compared to pedestrians acting under local congestion. This effect is caused by the larger personal space, the nonlocal congestion term (1.1) is an average over a bigger set hence allowing for higher densities in attractive areas. Higher densities will in turn allow for more effective emergency planning when designing for example escape routes. The numerical simulation in the end of this paper confirms this effect. The pedestrians are allowed to move freely, but the observed effect will become even more beneficial for a planner when introducing an environment for the pedestrians to interact with. In reality, crowd management is often done

by the strategic placement of obstacles such as pillars and walls. Furthermore, the pedestrians acting under nonlocal congestion travel at an overall lower congestion risk than their local congestion counterpart. This suggests that the crowd acting under nonlocal congestion could potentially move at a higher velocity than the crowd under local congestion which allows for faster and more successful evacuations.

In [12] the mean-field optimal control is characterized through a matching argument. This control is an approximate Nash equilibrium for the N -crowd. It is, for each pedestrian, the best response to the movement of the others. Furthermore, two N -crowds are considered where each pedestrian has crowd-specific preferences such as the target location and the congestion preference. The authors set up a mean-field game and show that it is equivalent to an optimal control problem in the case of a potential game model. In this paper, we look at the crowd from the bird's-eye view of an evacuation planner. We seek a 'simultaneous' optimal strategy for all the pedestrians involved in the crowd through a mean-field type optimal control approach for the single-crowd case and a mean-field type game approach for the multiple-crowd case.

The contributions of this paper are the following. We identify a model for the N -crowd that converges to the mean-field model proposed in [12]. This gives us insight into how the interaction between pedestrians in the N -crowd effects the mean-field model and reveals that the crowd of [12] is acting under local congestion. Our second contribution is a relaxation of the locality of the pedestrian model by allowing for interaction between pedestrians at a distance. Each pedestrian is given a personal space where it dislikes crowding, instead of interacting with other pedestrians only through collisions. This conceptual change is realistic since pedestrians do not need to be in physical contact to interact. As was discussed above, the suggested nonlocal congestion model allows for the following desirable features:

- Higher densities in attractive areas such as exits or escape routes where the pedestrians have to choose between higher congestion and penalty.
- Lower congestion risk, which implies a potential increase in pedestrian velocity allowing for faster exits and a larger flow of people, a very useful feature in the design of evacuation strategies.

Finally, we generalize the model to allow for an arbitrary number of interacting crowds and we treat it as a mean-field type game, in contrast to the mean-field game in [12]. We link the mean-field type game to an optimal control problem and prove a sufficient maximum principle. Treating the problem as a mean-field type game, the solution we find is not a Nash equilibrium but that of a central planner strategy.

The paper is organized as follows. After a short section of preliminaries, we consider the single-crowd case in Section 3. In Section 4, the multiple-crowd case is studied. The results derived in Section 3 generalize to an arbitrary finite number of interacting crowds and the solution to the mean-field type game is characterized by a sufficient maximum principle. An example that highlights the difference between local congestion and nonlocal congestion is solved numerically in Section 5. For the sake of clarity, all technical proofs are moved to an appendix.

2. Preliminaries

Given a general Polish space \mathcal{S} , let $\mathcal{P}(\mathcal{S})$ denote the space of probability measures on $\mathcal{B}(\mathcal{S})$. For an element $s \in \mathcal{S}$, the Dirac measure on s is an element of $\mathcal{P}(\mathcal{S})$ and will be denoted by δ_s . Let $\mathcal{P}(\mathcal{S})$ be equipped with the topology of weak convergence of probability measures. A metric that induces this topology is the bounded 1-Lipschitz metric,

$$d_{\mathcal{P}(\mathcal{S})}(\mu, \nu) := \|\mu - \nu\|_1 = \sup_{f \in L_1} \langle \mu, f \rangle - \langle \nu, f \rangle, \quad (2.1)$$

where L_1 is the set of real-valued functions on \mathcal{S} bounded by 1 and with Lipschitz coefficient 1. With this metric, $\mathcal{P}(\mathcal{S})$ is a Polish space. The space of probability measures on $\mathcal{B}(\mathcal{S})$ with finite second moments will be denoted by $\mathcal{P}_2(\mathcal{S})$,

$$\mathcal{P}_2(\mathcal{S}) := \left\{ \nu \in \mathcal{P}(\mathcal{S}) : \exists s_0 \in \mathcal{S} \text{ that satisfies } \int_{\mathcal{S}} d_{\mathcal{S}}(s, s_0)^2 \nu(ds) < \infty \right\}. \quad (2.2)$$

Equipped with the topology of weak convergence of measures and convergence of second moments, $\mathcal{P}_2(\mathcal{S})$ is a Polish space. A compatible complete metric is the square Wasserstein metric $d_{\mathcal{P}_2(\mathcal{S})}$,

$$d_{\mathcal{P}_2(\mathcal{S})}(\nu, \tilde{\nu}) := \left(\inf_{\substack{\alpha \in \mathcal{P}(\mathcal{S} \times \mathcal{S}) \\ \alpha_1 = \nu, \alpha_2 = \tilde{\nu}}} \int_{\mathcal{S} \times \mathcal{S}} d_{\mathcal{S}}(s, \tilde{s})^2 \alpha(ds, d\tilde{s}) \right)^{1/2} \quad (2.3)$$

where α_1 (α_2) denotes the first (second) marginal of α . From (2.3) we have the two following inequalities. For all $s_i, \tilde{s}_i \in \mathcal{S}$ and for all $N \in \mathbb{N}$,

$$d_{\mathcal{P}_2(\mathcal{S})}^2 \left(\frac{1}{N} \sum_{i=1}^N \delta_{s_i}, \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{s}_i} \right) \leq \frac{1}{N} \sum_{i=1}^N d_{\mathcal{S}}(s_i, \tilde{s}_i)^2. \quad (2.4)$$

For X with distribution ν and \tilde{X} with distribution $\tilde{\nu}$,

$$d_{\mathcal{P}_2(\mathcal{S})}^2(\nu, \tilde{\nu}) \leq \mathbb{E} \left[|X - \tilde{X}|^2 \right]. \quad (2.5)$$

Let $T > 0$ be a finite time horizon and let \mathbb{R}^d , $d \in \mathbb{N}$, be equipped with the Euclidean norm. Let \mathcal{M} and \mathcal{M}_2 be the spaces of continuous functions on $[0, T]$ with values in $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_2(\mathbb{R}^d)$ respectively,

$$\mathcal{M} := C([0, T]; \mathcal{P}(\mathbb{R}^d)), \quad \mathcal{M}_2 := C([0, T]; \mathcal{P}_2(\mathbb{R}^d)). \quad (2.6)$$

Equipped with the uniform metrics $d_{\mathcal{M}}$ and $d_{\mathcal{M}_2}$ given by

$$d_{\mathcal{M}}(m, m') := \sup_{t \in [0, T]} d_{\mathcal{P}(\mathbb{R}^d)}(m_t, m'_t), \quad d_{\mathcal{M}_2}(m, m') := \sup_{t \in [0, T]} d_{\mathcal{P}_2(\mathbb{R}^d)}(m_t, m'_t), \quad (2.7)$$

\mathcal{M} and \mathcal{M}_2 are Polish spaces. The mathematical results stated above can be found in [16, Chapter 2] and [8, Chapter 14].

Let A be a compact subset of \mathbb{R}^d . Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, denote by \mathcal{A} the set of A -valued \mathbb{F} -adapted processes such that

$$\mathbb{E} \left[\int_0^T |a_t|^2 dt \right] < \infty. \quad (2.8)$$

An element of \mathcal{A} will be called an *admissible control*. From the context, it will be clear which stochastic basis the notation \mathcal{A} is referring to.

Given a vector $x = (x^1, \dots, x^N)$ in the product space \mathcal{S}^N and an element $y \in \mathcal{S}$, we let

$$\begin{aligned} x^{-i} &:= (x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^N), \\ (y, x^{-i}) &:= (x^1, \dots, x^{i-1}, y, x^{i+1}, \dots, x^N). \end{aligned} \quad (2.9)$$

Furthermore, the law of any random quantity X will be denoted by $\mathcal{L}(X)$ and any index set of the form $\{1, \dots, N\}$ will be denoted by $\llbracket N \rrbracket$.

3. Single-crowd congestion

3.1. The N -crowd model

Let $(\Omega_N, \mathcal{F}^N, \mathbb{F}^N, \mathbb{P}_N)$ be a complete filtered probability space for each $N \in \mathbb{N}$. The filtration \mathbb{F}^N is right-continuous and augmented with \mathbb{P}_N -null sets. It carries the independent d -dimensional \mathbb{F}^N -Wiener processes $W^{1,N}, \dots, W^{N,N}$. Let, for each $i \in \llbracket N \rrbracket$, the \mathcal{F}_0^N -measurable \mathbb{R}^d -valued random variable $\xi^{i,N}$ be square-integrable and independent of $(W^{1,N}, \dots, W^{N,N})$. Given a vector of admissible controls, $\bar{a}^N = (a^{1,N}, \dots, a^{N,N}) \in \mathcal{A}^N$, consider the system

$$dX_t^{i,N} = b(t, X_t^{i,N}, a_t^{i,N})dt + \sigma(t, X_t^{i,N})dW_t^{i,N}, \quad X_0^{i,N} = \xi^{i,N}, \quad i \in \llbracket N \rrbracket. \quad (3.1)$$

We make the following assumptions on the coefficients b and σ in order to have a unique strong solution to (3.1).

- (A1) $b : [0, T] \times \mathbb{R}^d \times A \rightarrow \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ are continuous in all arguments.
- (A2) For all $x_1, x_2 \in \mathbb{R}^d$ and $a_1, a_2 \in A$, there exists a constant $K > 0$ independent of (t, x_1, x_2, a_1, a_2) such that

$$\begin{aligned} |b(t, x_1, a_1) - b(t, x_2, a_2)| &\leq K(|x_1 - x_2| + |a_1 - a_2|), \\ |\sigma(t, x_1) - \sigma(t, x_2)| &\leq K|x_1 - x_2|, \\ |b(t, x_1, a_1)| + |\sigma(t, x_1)| &\leq K(1 + |x_1| + |a_1|). \end{aligned}$$

Proposition 3.1. *Under (A1) and (A2) the system (3.1) has a unique strong solution in the sense that*

$$X_0^{i,N} = \xi^{i,N}, \quad (3.2)$$

$$\int_0^t |b(s, X_s^{i,N}, a_s^{i,N})| + |\sigma(s, X_s^{i,N})|^2 ds < \infty, \quad t \in [0, T], \quad \mathbb{P} - a.s. \quad (3.3)$$

$$X_t^{i,N} = \xi^{i,N} + \int_0^t b(s, X_s^{i,N}, a_s^{i,N})ds + \int_0^t \sigma(s, X_s^{i,N})dW_s^{i,N}, \quad t \in [0, T]. \quad (3.4)$$

Furthermore, the strong solution $X^{i,N}$ satisfies the estimate

$$\mathbb{E}_N \left[\sup_{s \in [0,t]} |X_s^{i,N}|^2 \right] \leq K_t (1 + \mathbb{E}_N [|\xi^{i,N}|^2]) \quad (3.5)$$

for all $t \in [0, T]$, for all $i \in \llbracket N \rrbracket$ and for some positive constant K_t depending only on t .

Proof. A proof can be found in [20, Chapter 1, Theorem 6.16]. Note that K_t is independent of $a^{i,N}$ by compactness of A . \square

The process $X^{i,N}$ models the motion of a pedestrian in the N -crowd who partially controls its velocity through the control $a^{i,N}$. Since its control is adapted to the full filtration \mathbb{F}^N , the model allows for the pedestrian to take every movement in the crowd into account. Its motion is also influenced by external forces, such as the random disturbance driven by $W^{i,N}$. The motion of the pedestrian may be modeled more generally than above by introducing an explicit weak interaction in the drift [7], such as

$$dX_t^{i,N} = \frac{1}{N} \sum_{j=1}^N \tilde{b}(t, X_t^{i,N}, a_t^{i,N}, X_t^{j,N}) dt + \sigma(t, X_t^{i,N}) dW_t^{i,N}. \quad (3.6)$$

It is also possible to let a common disturbance effect all pedestrians [10], to model for example evacuations during an earthquake, a fire, a tsunami etc.

Pedestrian i evaluates the state of the N -crowd, given by the control vector $\bar{a}^N = (a^{1,N}, \dots, a^{N,N})$, according to its individual congestion risk measure $J_r^{i,N}$,

$$J_r^{i,N}(\bar{a}^N) := \mathbb{E}_N \left[\int_0^T \left(\frac{1}{2} |a_t^{i,N}|^2 + \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y) \mu_t^{-i,N}(dy) \right) dt + \Psi(X_T^{i,N}) \right]. \quad (3.7)$$

where $\mu_t^{-i,N}$ is the empirical measure of $X^{-i,N}$,

$$\mu_t^{-i,N} := \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \delta_{X_t^{j,N}}, \quad (3.8)$$

and $X^{1,N}, \dots, X^{N,N}$ is the solution of (3.1) given \bar{a}^N . The region where crowding has an influence on the pedestrian's choice of control, its 'personal space', is ideally modeled by a normalized indicator function,

$$\mathbb{I}_r(x) := \begin{cases} \text{Vol}(B_r)^{-1}, & x \in B_r, \\ 0, & x \notin B_r, \end{cases} \quad (3.9)$$

where $B_r \subset \mathbb{R}^d$ is the ball with radius $r > 0$ centered at the origin and $\text{Vol}(B_r)$ is its volume. The term

$$\int_{\mathbb{R}^d} \mathbb{I}_r(X_t^{i,N} - y) \mu_t^{-i,N}(dy) \quad (3.10)$$

then represents the number of pedestrians around $X_t^{i,N}$ (except itself) within a distance less than r at time t [18]. To simplify the calculations we will use a smoothed version of \mathbb{I}_r . Let γ_δ be a mollifier,

$$\gamma_\delta(x) := \gamma(x/\delta)/\delta, \quad (3.11)$$

where γ is a smooth symmetric probability density with compact support. For a fixed $\delta > 0$, we define

$$\phi_r(x) := \gamma_\delta * \mathbb{I}_r(x). \quad (3.12)$$

For convergence estimates later in this section, we assume that the final cost Ψ satisfies the following condition.

(A3) For all $x_1, x_2 \in \mathbb{R}^d$ there exists a constant $K > 0$ independent of (x_1, x_2) such that

$$|\Psi(x_1) - \Psi(x_2)| \leq K|x_1 - x_2|.$$

The interpretation of the congestion risk measure is the following. The first term penalizes large controls (energy waste) whereas the second term penalizes congestion within a personal space. The final cost penalizes deviations from specific target regions at time T . Typically the final cost takes large values everywhere except in areas where the pedestrians want to end up, places like meeting points, evacuation doors, etc.

3.2. The mean-field type control problem

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space such that the filtration is right continuous and augmented with \mathbb{P} -null sets. Let \mathbb{F} carry a Wiener process W and let ξ be an \mathcal{F}_0 -measurable and square-integrable \mathbb{R}^d -valued random variable independent of W . Given a control $a \in \mathcal{A}$, the mean-field type dynamics is

$$dX_t = b(t, X_t, a_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = \xi. \quad (3.13)$$

By Proposition 3.1 there exists a unique strong solution to (3.13). The mean-field type congestion risk measure is given by

$$J_r(a) = \mathbb{E} \left[\int_0^T \frac{1}{2} |a_t|^2 dt + \int_{\mathbb{R}^d} \phi_r(X_t - y) \mu_{X_t}(dy) dt + \Psi(X_T) \right]. \quad (3.14)$$

where μ_{X_t} is the distribution of X_t .

Remark 3.1. The difference between a mean-field type control and a mean-field game is that in general mean-field games can be reduced to a standard control problem and an equilibrium while a mean-field type control problem is a nonstandard control problem ([2],[4]). The matching procedure to find the fixed point (equilibrium) for a mean-field game is pedagogically described as follows ([7],[13]).

- (i) Fix a deterministic function $\mu_t : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$.

(ii) Solve the standard stochastic control problem

$$\hat{a} = \operatorname{argmin}_{a \in \mathcal{A}} \mathbb{E} \left[\int_0^T \frac{1}{2} |a_t|^2 + \int_{\mathbb{R}^d} \phi_r(X_t - y) \mu_t(dy) dt + \Psi(X_T) \right], \quad (3.15)$$

where X is the dynamics corresponding to a .

(iii) Determine the function $\hat{\mu}_t : [0, T] \rightarrow \mathcal{P}_2(\mathbb{R}^d)$ such that $\hat{\mu}_t = \mathcal{L}(\hat{X}_t)$ for all $t \in [0, T]$ where \hat{X} is the dynamics corresponding to the optimal control \hat{a} .

In the mean-field type control setting, the measure-valued process $(\mu_{X_t}; t \in [0, T])$ is not considered to be a separate variable but is given by the input of the congestion risk. This non-linearity makes the control problem nonstandard.

3.3. Convergence of the processes

We make the following assumptions about the initial data $\xi^{1,N}, \dots, \xi^{N,N}$.

(B1) $\sup_{N \in \mathbb{N}} \mathbb{E}_N \left[\frac{1}{N} \sum_{i=1}^N |\xi^{i,N}|^2 \right] < \infty$ for all $i \in \llbracket N \rrbracket$.

(B2) $(\xi^{1,N}, \dots, \xi^{N,N})$ is exchangeable for all $N \in \mathbb{N}$.

(B3) $\lim_{N \rightarrow \infty} \mathcal{L} \left(\frac{1}{N} \sum_{i=1}^N \delta_{\xi^{i,N}} \right) = \delta_{\mu_0}$ in $\mathcal{P}(\mathcal{P}_2(\mathbb{R}^d))$.

Under assumption (B1) and (B2), the sequence $(\xi^{i,N})_{N \in \mathbb{N}}$ is tight for all $i \in \llbracket N \rrbracket$ [5]. Under assumptions (B1), (B2) and (B3) $\xi^{i,N}$ converges weakly in $\mathcal{P}_2(\mathbb{R}^d)$ to a μ_0 -distributed random variable, from now on denoted by ξ , for all $i \in \llbracket N \rrbracket$ [5]. We make the following assumption about the controls.

(B4) The controls are of feedback form, $a_t^{i,N}(\omega) = a^N(t, X_t^{i,N}(\omega))$, where each a^N is an A -valued deterministic function and a^N converge uniformly to a as $N \rightarrow \infty$. Furthermore,

$$\sup_{N \in \mathbb{N}} \mathbb{E}_N \left[\int_0^T |a^N(t, X_t^{i,N})|^2 \right] < \infty, \quad \forall i \in \llbracket N \rrbracket. \quad (3.16)$$

Remark 3.2. Assumption (B4) implies that, while the paths of pedestrians in the N -crowd may differ, they are outcomes from a symmetric joint probability distribution. By exchangeability of $(\xi^{i,N}, W^{i,N})_{i=1}^N$,

$$(a^N(t, X_t^{i,N}))_{i=1}^N \stackrel{d}{=} (a^N(t, X_t^{\pi(i),N}))_{i=1}^N \quad (3.17)$$

for all permutations π of $\llbracket N \rrbracket$ and the interpretation is that we cannot distinguish pedestrians in the crowd. The pedestrians are anonymous.

Let $X^{1,N}, \dots, X^{N,N}$ be the solution of (3.1) given a^N and let μ^N denote the empirical measure process of these states,

$$\mu^N = \frac{1}{N} \sum_{i=1}^N \delta_{X^{i,N}}. \quad (3.18)$$

Proposition 3.2. *The collection of measures $\{\mathcal{L}(\mu^N), N \in \mathbb{N}\}$ in $\mathcal{P}(\mathcal{M}_2)$ is tight.*

Proof. The empirical measures are elements of \mathcal{M}_2 by Proposition 3.1 together with (B1) and (B2). The proof of tightness is found in Appendix 6.1. \square

Recall that a sequence $\{X_n\}$ of random variables converges weakly to X in a Polish space if and only if $\{X_n\}$ is tight and every convergent subsequence of $\{X_n\}$ converges to X . The tightness of the empirical measures implies that along a converging subsequence, μ^N converges weakly to the measure-valued process μ that satisfies the integral equation

$$\langle \mu_t, f \rangle = \langle \mu_0, f \rangle + \int_0^t \left\langle \mu_s, b(s, \cdot, a(s, \cdot)) \cdot \nabla f + \frac{1}{2} \sigma(s, \cdot) \Delta f \right\rangle ds, \quad \forall f \in C_b^2(\mathbb{R}^d). \quad (3.19)$$

Since the strong solution of (3.13) is unique, the weak solution is also unique [19] which is equivalent to uniqueness of solutions to (3.19) [9]. We have the following result.

Theorem 3.1. *Let X^i , $i \in \mathbb{N}$, be independent copies of the strong solution of (3.13). Under assumptions (A1), (A2) and (B1)-(B4) the state process $X^{i,N}$ for a pedestrian in the N -crowd converges weakly to X^i as $N \rightarrow \infty$.*

Proof. Applying the propagation of chaos theorem of Sznitman [17], the theorem follows by the weak convergence of μ^N to the deterministic measure μ . \square

3.4. Convergence of the congestion risk measure

From the previous section we know that $X^{i,N}$, the strong solution of (3.1), converges weakly to X , the strong solution of (3.13), and we know that μ_t^N converges weakly to μ_{X_t} . Applying (2.4), we get that

$$d_{\mathcal{P}_2(\mathbb{R}^d)}(\mu_t^{-i,N}, \mu_t^N) \leq 2/N, \quad (3.20)$$

so $\mu_t^{-i,N}$ converges weakly to μ_{X_t} as well. By Skorokhod's Representation Theorem [8, Theorem 3.30] we can represent (up to distribution) all the random variables mentioned above in a common probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where they converge $\tilde{\mathbb{P}}$ -almost surely. This allows us to write

$$\begin{aligned} |J_r^{i,N}(a^N) - J_r(a)| &\leq \tilde{\mathbb{E}} \left[\int_0^T \left| \frac{1}{2} |a^N(t, X_t^{i,N})|^2 - \frac{1}{2} |a(t, X_t)|^2 \right| + \left| \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y) \mu_t^{-i,N}(dy) \right. \right. \\ &\quad \left. \left. - \int_{\mathbb{R}^d} \phi_r(X_t - y) \mu_{X_t}(dy) \right| dt + \left| \Psi(X_T^{i,N}) - \Psi(X_T) \right| \right], \end{aligned} \quad (3.21)$$

where the expectation is taken under $\tilde{\mathbb{P}}$. By compactness of A , the Continuous Mapping Theorem, (B4) and Dominated Convergence we have

$$\lim_{N \rightarrow \infty} \tilde{\mathbb{E}} \left[\int_0^T \left| \frac{1}{2} |a^N(t, X_t^{i,N})|^2 - \frac{1}{2} |a(t, X_t)|^2 \right| \right] = 0. \quad (3.22)$$

By (A3), Proposition 3.1 and Dominated Convergence,

$$\lim_{N \rightarrow \infty} \widetilde{\mathbb{E}} \left[\left| \Psi(X_T^{i,N}) - \Psi(X_T) \right| \right] = 0. \quad (3.23)$$

Note that

$$\begin{aligned} & \widetilde{\mathbb{E}} \left[\int_0^T \left| \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y) \mu_t^{-i,N}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t - y) \mu_{X_t}(dy) \right| dt \right] \\ & \leq \widetilde{\mathbb{E}} \left[\int_0^T \left| \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y) \mu_t^{-i,N}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y) \mu_{X_t}(dy) \right| dt \right] \\ & \quad + \widetilde{\mathbb{E}} \left[\int_0^T \left| \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y) \mu_{X_t}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t - y) \mu_{X_t}(dy) \right| dt \right]. \end{aligned} \quad (3.24)$$

The first term on the right hand side tends to zero as $N \rightarrow \infty$ by the definition of weak convergence while the second tends to zero as $N \rightarrow \infty$ by the Continuous Mapping Theorem and Dominated Convergence. We have proved the following result.

Theorem 3.2. *Let $a \in \mathcal{A}$ and let $a^N = (a, \dots, a) \in \mathcal{A}^N$. Then*

$$J_r^{i,N}(a^N) = J_r(a) + \varepsilon_N, \quad (3.25)$$

where $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$.

3.5. Solutions to NP-1 and MFT-1

The notion of solutions of the the N -crowd model (NP-1) and the mean-field type control model (MFT-1) for single-crowd congestion will now be defined.

Definition 3.1 (Solution to NP-1). *Let $\hat{a}^N = (\hat{a}, \dots, \hat{a}) \in \mathcal{A}^N$ for some fixed $\hat{a} \in \mathcal{A}$ and let $a^N = (a, \dots, a) \in \mathcal{A}^N$ for an arbitrary strategy $a \in \mathcal{A}$. Then \hat{a}^N is a solution to NP-1 if*

$$J_r^{i,N}(\hat{a}^N) \leq J_r^{i,N}(a^N), \quad \forall a \in \mathcal{A}, \quad \forall i \in \llbracket N \rrbracket. \quad (3.26)$$

If, for a given $\varepsilon > 0$, \hat{a} satisfies

$$J_r^{i,N}(\hat{a}^N) \leq J_r^{i,N}(a^N) + \varepsilon, \quad \forall a \in \mathcal{A}, \quad \forall i \in \llbracket N \rrbracket, \quad (3.27)$$

then \hat{a}^N is an ε -solution to NP-1.

Definition 3.2 (Solution to MFT-1). *If $\hat{a} \in \mathcal{A}$ satisfies*

$$J_r(\hat{a}) \leq J_r(a), \quad \forall a \in \mathcal{A}, \quad (3.28)$$

then \hat{a} is a solution to MFT-1.

The following result motivates the use of MFT-1 as an approximation to NP-1. It confirms that we can construct an approximate solution to NP-1 using a solution to MFT-1 and the approximation improves as the size of the crowd increases.

Theorem 3.3. *If \hat{a} solves MFT-1, then $\hat{a}^N = (\hat{a}, \dots, \hat{a})$ is a $2\varepsilon_N$ -solution to NP-1 among feedback strategies.*

Proof. We apply (3.25) and the definition of a solution to MFT-1. Let $a^N = (a, \dots, a) \in \mathcal{A}^N$ for some $a \in \mathcal{A}$. It holds that

$$\begin{aligned} J_r^{i,N}(a^N) &= J_r(a) + \varepsilon_N \\ &\geq J_r(\hat{a}) + \varepsilon_N \\ &= J_r^{i,N}(\hat{a}^N) + 2\varepsilon_N. \end{aligned} \tag{3.29}$$

□

Remark 3.3. It is known that the solution of a mean-field game corresponds to an approximate Nash equilibrium for NP-1 ([7],[13]). To the best of our knowledge, this has not been shown to be true for solutions to mean-field type control problems. Theorem 3.3 has the following interpretation. A mean-field type optimal control induces an approximate solution for the N -crowd if the crowd consists homogeneous pedestrians and thus a representative pedestrian determines the control of all. This was actually already visible in Theorem 3.2.

3.6. Deterministic version of MFT-1

We want to present results in a setting similar to [12] to highlight the differences between the models. To do this, we make the assumption that μ_{X_t} has a density $m_X(t, \cdot)$ for all $t \in [0, T]$. An example of sufficient conditions for the existence is bounded drift and diffusion [15]. Under this assumption, we may rewrite (3.13)-(3.14) into a deterministic problem for m_X . Furthermore, an admissible control can not be stochastic in the deterministic problem formulation. The full stochastic problem will be analyzed in future work. We have a new definition of an admissible control.

Definition 3.3 (\mathcal{A}_d). *A square-integrable deterministic function $a : [0, T] \times \mathbb{R}^d \rightarrow A$ will be called an admissible control for the deterministic problem and the set of such functions is denoted by \mathcal{A}_d .*

By (3.19) the density m_X satisfies

$$\begin{aligned} \int_{\mathbb{R}^d} f(x) m_X(t, x) dx &= \int_{\mathbb{R}^d} f(x) m_X(0, x) dx + \int_0^t \int_{\mathbb{R}^d} \left(b(s, x, a(s, x)) \cdot \nabla f(x) \right. \\ &\quad \left. + \frac{1}{2} \text{Tr} [\sigma \sigma^T(s, x) \Delta f(x)] \right) m_X(s, x) ds dx, \quad \forall f \in C_b^2(\mathbb{R}^d), \quad \forall t \in [0, T]. \end{aligned} \tag{3.30}$$

Hence it is a weak solution to the Fokker-Planck equation,

$$\begin{cases} \frac{\partial m_X}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2 \sigma \sigma^T m_X(t, x)] - \nabla \cdot (b(t, x, a(t, x)) m_X(t, x)), \\ m_X(0, x) = \text{density of } \mu_0. \end{cases} \tag{3.31}$$

We arrive to a deterministic version of MFT-1 (dMFT-1),

$$\begin{aligned} & \text{minimize} && J_r^{\det}(a) \\ & \text{subject to} && \begin{cases} \frac{\partial m_X}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2 \sigma \sigma^T m_X(t, x)] - \nabla \cdot (b(t, x, a(t, x))m(t, x)), \\ m(0, x) = \text{density of } \mu_0, \\ a \in \mathcal{A}_d, \end{cases} \end{aligned} \quad (3.32)$$

where

$$\begin{aligned} J_r^{\det}(a) := & \int_{\mathbb{R}^d} \left[\int_0^T \left(\frac{1}{2} |a(t, x)|^2 m_X(t, x) \right. \right. \\ & \left. \left. + \left(\int_{\mathbb{R}^d} \phi_r(x - y) m_X(t, y) dy \right) m_X(t, x) \right) dt + \Psi(x) m(T, x) \right] dx. \end{aligned} \quad (3.33)$$

Remark 3.4. Note that ϕ_r converges weakly to δ_0 as $r \rightarrow 0$. In this limit, the congestion risk measure tends to

$$J_0^{\det}(a) = \int_{\mathbb{R}^d} \int_0^T \frac{1}{2} |a(t, x)|^2 m_X(t, x) + m_X(t, x)^2 dt + \Psi(x) m(T, x) dx, \quad (3.34)$$

which is exactly the risk analyzed in the pedestrian crowd model of [12]! Clearly the deterministic model of [12] corresponds to a situation where the pedestrian will only react to how likely it is to 'bump' into other pedestrians. In the case of positive r , a pedestrian is effected by crowding within a personal space of nonzero range and reacts to the level of the density within this range. This is the clear distinction between local and nonlocal congestion.

4. Multiple-crowd congestion

4.1. The N -crowd model

The multiple-crowd case corresponds to a situation where the terminal goal and/or opinion on congestion differ between subsets of pedestrians. This inhomogeneity is introduced in the congestion risk measure.

Let $(\Omega_N, \mathcal{F}^N, \mathbb{F}^N, \mathbb{P}_N)$ be a complete filtered probability space for each $N \in \mathbb{N}$. The filtration \mathbb{F}^N is right-continuous and augmented with \mathbb{P}^N -null sets. It carries the independent d -dimensional \mathbb{F}^N -Wiener processes $\mathbf{W} = ((W^{i,j,N})_{i=1}^N)_{j=1}^M$. Let, for each $i \in \llbracket N \rrbracket$ and for each $j \in \llbracket M \rrbracket$, the \mathcal{F}_0^N -measurable \mathbb{R}^d -valued random variable $\xi^{i,j,N}$ be square-integrable and independent \mathbf{W} . Given a vector of admissible controls, $\mathbf{a} = ((a^{i,j,N})_{i=1}^N)_{j=1}^M \in A^{NM}$, consider the system

$$\begin{cases} dX_t^{i,j,N} = b(t, X_t^{i,j,N}, a_t^{i,j,N}) dt + \sigma(t, X_t^{i,j,N}) dW_t^{i,j,N}, \\ X_0^{i,j,N} = \xi^{i,j,N}, \quad i \in \llbracket N \rrbracket, j \in \llbracket M \rrbracket. \end{cases} \quad (4.1)$$

In view of Proposition 3.1 there exists a unique strong solution to (4.1). Pedestrian i in crowd j evaluates \mathbf{a} according to its individual congestion risk measure

$$J_{r,\Lambda}^{i,j,N}(\mathbf{a}) := \mathbb{E}_N \left[\int_0^T \frac{1}{2} |a^{i,j,N}|^2 + \int_{\mathbb{R}^d} \phi_r(X_t^{i,j,N} - y) \tilde{\nu}_{t,\Lambda}^{j,N}(dy) dt + \Psi_j(X_T^{i,j,N}) \right], \quad (4.2)$$

where

$$\tilde{\nu}_{t,\Lambda}^{j,N} := \sum_{k=1}^M \lambda_{jk} \frac{1}{N} \sum_{l=1}^N \delta_{X_t^{l,k,N}}, \quad (4.3)$$

λ_{jk} are bounded and non-negative real numbers and $\Lambda = (\lambda_{jk})_{jk}$. The λ_{jk} are interpreted as congestion weights. Indeed the weights λ_{jk} quantify the congestion preferences in the model. If λ_{jk} is high, pedestrians in crowd j pay a high price for being close to pedestrians in crowd k . If λ_{jk} is zero, pedestrians in crowd j are indifferent to the positioning of pedestrians in crowd k . Note that if $\lambda_{jk} = 1$ for $j = k$ and 0 otherwise, the crowds are disconnected in the sense that there is no interaction between pedestrians in different crowds and each crowd is modeled as an NP-1.

Assumptions (A1) and (A2) are still valid, but we have to modify (A3) to the new setting. Assume that the final costs Ψ_j , $j \in \llbracket M \rrbracket$, satisfy the following condition.

(A3*) For all $x_1, x_2 \in \mathbb{R}^d$ there exists a constant $K > 0$ independent of (x_1, x_2) such that

$$|\Psi_j(x_1) - \Psi_j(x_2)| \leq K|x_1 - x_2|, \quad \forall j \in \llbracket M \rrbracket. \quad (4.4)$$

4.2. The mean-field type model

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space such that the filtration satisfies the usual conditions and carries a d -dimensional independent \mathbb{F} -Wiener processes W^1, \dots, W^M . Let, for each $j \in \llbracket M \rrbracket$, ξ^j be an \mathcal{F}_0 -measurable and square-integrable random variable independent of (W^1, \dots, W^M) . Given a vector of controls $\bar{a}^M = (a^1, \dots, a^M)$, the mean-field type dynamics are given by

$$dX_t^j = b(t, X_t^j, a_t^j)dt + \sigma(t, X_t^j)dW_t^j, \quad X_0^j = \xi^j, \quad j \in \llbracket M \rrbracket. \quad (4.5)$$

There exists a unique strong solution to (4.5) by Proposition 3.1. The mean-field type congestion risk for crowd $j \in \llbracket M \rrbracket$ is given by

$$J_{r,\Lambda}^j(\bar{a}^M) := \mathbb{E} \left[\int_0^T \frac{1}{2} |a^j|^2 + \int_{\mathbb{R}^d} \phi_r(X_t^j - y) \nu_{t,\Lambda}^j(dy) dt + \Psi_j(X_T^j) \right], \quad (4.6)$$

where $\nu_{t,\Lambda}^j := \sum_{k=1}^M \lambda_{jk} \mu_{X_t^k}$.

4.3. Solutions of NP-M and MFT-M

The convergence results for the single-crowd case generalizes to multiple crowds under the following assumptions.

- (C1) $\sup_{N \in \mathbb{N}} \mathbb{E}_N \left[\frac{1}{N} \sum_{i=1}^N |\xi^{i,j,N}|^2 \right] < \infty$ for all $j \in \llbracket M \rrbracket$.
 (C2) $(\xi^{1,j,N}, \dots, \xi^{N,j,N})$ is exchangeable for all $j \in \llbracket M \rrbracket$.
 (C3) $\lim_{N \rightarrow \infty} \mathcal{L} \left(\frac{1}{N} \sum_{i=1}^N \xi^{i,j,N} \right) = \delta_{\mu_j^0}$ in $\mathcal{P}(\mathcal{P}_2(\mathbb{R}^d))$ for all $j \in \llbracket M \rrbracket$.
 (C4) The controls are of feedback form, $a_t^{i,j,N}(\omega) = a^{j,N}(t, X_t^{i,j,N}(\omega))$ where each $a^{j,N}$ is a deterministic A -valued function and $a^{j,N}$ converge uniformly to a^j as $N \rightarrow \infty$. Furthermore,

$$\sup_{N \in \mathbb{N}} \mathbb{E}_N \left[\int_0^T |a^{j,N}(t, X_t^{i,j,N})|^2 \right] < \infty, \quad \forall i \in \llbracket N \rrbracket, \forall j \in \llbracket M \rrbracket. \quad (4.7)$$

Under (A1),(A2),(A3*) and (C1)-(C4) the results from Section 3.3 and Section 3.4 generalize to M crowds without any complications. Solutions to the N -crowd model (NP-M) and the mean-field type model (MFT-M) for multiple-crowd congestion are defined below.

Definition 4.1 (Solution to NP-M). *For each $j \in \llbracket M \rrbracket$, fix a control $\hat{a}^j \in \mathcal{A}$ and let $(\hat{a}^j)^N = (\hat{a}^j, \dots, \hat{a}^j) \in \mathcal{A}^N$. For an arbitrary control $a^j \in \mathcal{A}$, let $(a^j)^N = (a^j, \dots, a^j) \in \mathcal{A}^N$. The control vector $((\hat{a}^1)^N, \dots, (\hat{a}^M)^N)$ is a solution to NP-M if*

$$J_{r,\Lambda}^{i,j,N}((\hat{a}^1)^N, \dots, (\hat{a}^M)^N) \leq J_{r,\Lambda}^{i,j,N}((a^j)^N, (\hat{a}^{-j})^N), \quad \forall a^j \in \mathcal{A}, \forall j \in \llbracket M \rrbracket. \quad (4.8)$$

If, for a given $\varepsilon > 0$, the control vector $((\hat{a}^1)^N, \dots, (\hat{a}^M)^N)$ satisfies

$$J_{r,\Lambda}^{i,j,N}((\hat{a}^1)^N, \dots, (\hat{a}^M)^N) \leq J_{r,\Lambda}^{i,j,N}((a^j)^N, (\hat{a}^{-j})^N) + \varepsilon, \quad \forall a^j \in \mathcal{A}, \forall j \in \llbracket M \rrbracket, \quad (4.9)$$

then $((\hat{a}^1)^N, \dots, (\hat{a}^M)^N)$ is an ε -solution to MFT-M.

Definition 4.2 (Solution to MFT-M). *The control vector $\hat{a}^M = (\hat{a}^{1,M}, \dots, \hat{a}^{M,M}) \in \mathcal{A}^M$ is a solution to MFT-M if*

$$J_{r,\Lambda}^j(\hat{a}^M) \leq J_{r,\Lambda}^j(a, \hat{a}^{-j,M}), \quad \forall a \in \mathcal{A}, \forall j \in \llbracket M \rrbracket. \quad (4.10)$$

Remark 4.1. There is a fundamental difference between the definition of solutions in the single-crowd case and in the multiple-crowd case. The latter is a Nash equilibrium while the former is an optimal control. So, what has changed? We still have anonymity between pedestrians within a crowd but the vector of all controls used in the multiple-crowd case, $((a^{j,N}(t, X_t^{i,j,N}))_{i=1}^N)_{j=1}^M$ for NP-M and $(a^j(t, X_t^j))_{j=1}^M$ for MFT-M, is not exchangeable (cf. (3.17)). From our point of view, we may distinguish between two pedestrians from different crowds and hence the pedestrians are not anonymous anymore. Thus, it makes sense to look at a game problem between the crowds.

The approximation result in Theorem 3.3 generalizes to the multiple-crowd case.

Theorem 4.1. *Assume that \hat{a}^M is a solution to MFT-M. Then $((\hat{a}^{1,M})^N, \dots, (\hat{a}^{M,M})^N)$ is an $2\varepsilon_N$ -solution to NP-M.*

Proof. The proof follows exactly the same steps as the proof of Theorem 3.3. \square

Finally, under the assumption that $\mu_{X_t^j}$ admits a density $m_{X^j}(t, \cdot)$ we rewrite MFT-M into a deterministic problem (dMFT-M) where we optimize over the set \mathcal{A}_d (cf. Definition 3.3). The solution to dMFT-M is defined below.

Definition 4.3 (Solution to dMFT-M). A control vector $\hat{a} = (\hat{a}^1, \dots, \hat{a}^M) \in \mathcal{A}_d^M$ solves dMFT-M if it satisfies

$$J_{r,\Lambda}^{j,\det}(\hat{a}) \leq J_{r,\Lambda}^{j,\det}(a, \hat{a}^{-j}), \quad \forall a \in \mathcal{A}_d, \quad \forall j \in \llbracket M \rrbracket, \quad (4.11)$$

where, for $a = (a^1, \dots, a^M) \in \mathcal{A}_d^M$,

$$\begin{aligned} J_{r,\Lambda}^{j,\det}(a) := & \int_{\mathbb{R}^d} \left[\int_0^T \left(\frac{1}{2} |a^j(t, x)|^2 m_j(t, x) \right. \right. \\ & \left. \left. + \sum_{k=1}^M \lambda_{jk} \left(\int_{\mathbb{R}^d} \phi_r(x-y) m_k(t, y) dy \right) m_j(t, x) \right) dt + \Psi_j(x) m_j(T, x) \right] dx \end{aligned} \quad (4.12)$$

and m_j solves (in a weak sense)

$$\begin{cases} \frac{\partial m_j}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2 (\sigma \sigma^T m_j)(t, x)] - \nabla \cdot (b(t, x, a^j) m_j(t, x)), \\ m_j(0, t) = m_{0,j}(x), \text{ the density of } \mu_0^j. \end{cases} \quad (4.13)$$

Remark 4.2. In the limit $r \rightarrow 0$ the congestion risk measure is

$$\begin{aligned} J_{0,\Lambda}^{j,\det}(a) = & \int_{\mathbb{R}^d} \left[\int_0^T \left(\frac{1}{2} |a(t, x)|^2 m_j(t, x) \right. \right. \\ & \left. \left. + \sum_{k=1}^M \lambda_{j,k} m_k(t, x) m_j(t, x) \right) dt + \Psi_j(x) m_j(T, x) \right] dx. \end{aligned} \quad (4.14)$$

The interpretation is the same as in the single-crowd model, when $r \rightarrow 0$ the personal space of the pedestrians shrink to a singleton and only collisions have an impact on the choice of control. Note that (4.14) with parameters $M = 2$, $\lambda_{11} = \lambda_{22} = 1$ and $\lambda_{12} = \lambda_{21} = \lambda$ is exactly the cost that appears in [12].

4.4. An optimal control problem equivalent to dMFT-M, $r = 0$.

In this section we introduce an optimal control problem which is shown to have the same solution as dMFT-M. Instead of solving the game problem, we may solve an optimal control problem which is done by proving a Pontryagin-type Maximum Principle.

To ease notation, let

$$\begin{aligned} |a(t, x)|^2 &:= (|a^1(t, x)|^2, \dots, |a^M(t, x)|^2), \\ m(t, x) &:= (m_1(t, x), \dots, m_M(t, x)), \\ \Psi(x) &:= (\Psi_1(x), \dots, \Psi_M(x)). \end{aligned} \quad (4.15)$$

Consider the following optimization problem,

$$\begin{aligned} & \text{minimize} \quad J_{0,\Lambda}(a), \\ & \text{subject to} \quad \begin{cases} \frac{\partial m_j}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2 (\sigma \sigma^T m_j)(t, x)] - \nabla \cdot (b(t, x, a^j(t, x)) m_j(t, x)), \\ m_j(0, x) = \text{density of } \mu_0^j, \end{cases} \quad (OC_0) \\ & \quad a \in \mathcal{A}_d^M \end{aligned}$$

where

$$J_{0,\bar{\Lambda}}(a) := \int_{\mathbb{R}^d} \left[\int_0^T \left(\frac{1}{2} |a(t,x)|^2 \cdot m(t,x) + m(t,x)^T \bar{\Lambda} m(t,x) \right) dt + \Psi(x) \cdot m(T,x) \right] dx \quad (4.16)$$

and $\bar{\Lambda} := \frac{1}{2}(\Lambda + \text{diag}(\Lambda))$. The optimal control problem (OC_0) can be interpreted as a minimization of a joint congestion risk measure, $J_{0,\bar{\Lambda}}$, for the crowds. We anticipate that the optimal control problem (OC_0) will give us a solution to dMFT-M. This conjecture is proven in the results that now follow and $\bar{\Lambda}$ naturally appears as the correct weight matrix for the joint cost in order for the conjecture to hold. The next proposition is the first link between dMFT-M and (OC_0) .

Proposition 4.1. *If $\hat{a} \in \mathcal{A}_d^M$ solves (OC_0) then \hat{a} is a solution to dMFT-M when $r = 0$.*

Proof. The proof is deferred to Appendix 6.2 □

An optimal control to (OC_0) will now be characterized through a sufficient maximum principle. Let

$$f(t, x, a, m) := \frac{1}{2} |a(t, x)|^2 \cdot m(t, x) + m(t, x)^T \bar{\Lambda} m(t, x), \quad (4.17)$$

$$g(x, m) := \Psi(x) \cdot m(T, x), \quad (4.18)$$

and let, with some abuse of notation,

$$\begin{aligned} \text{Tr} [\sigma \sigma^T \nabla^2 p(t, x)] &:= (\text{Tr} [\sigma \sigma^T \nabla^2 p_1(t, x)], \dots, \text{Tr} [\sigma \sigma^T \nabla^2 p_M(t, x)]), \\ \text{Tr} [\nabla^2 (\sigma \sigma^T m)(t, x)] &:= (\text{Tr} [\nabla^2 (\sigma \sigma^T m_1)(t, x)], \dots, \text{Tr} [\nabla^2 (\sigma \sigma^T m_M)(t, x)]). \end{aligned} \quad (4.19)$$

Theorem 4.2 (Sufficient maximum principle for (OC_0)). *Let $\hat{a} \in \mathcal{A}_d^M$, let*

$$H(t, x, a, m, p) := f(t, x, a, m) + \sum_{j=1}^M b(t, x, a^j(t, x)) m_j(t, x) \cdot \nabla p_j(t, x) \quad (4.20)$$

and let p solve the adjoint equation

$$\begin{cases} \frac{\partial p}{\partial t}(t, x) = - \left(D_m \hat{H}(t, x) + \frac{1}{2} \text{Tr} [\sigma \sigma^T \nabla^2 p(t, x)] \right), \\ p(T, x) = D_m \hat{g}(x), \end{cases} \quad (4.21)$$

where $D_m \hat{H}$ and $D_m \hat{g}$ are the Gâteaux derivatives with respect to m in the point (\hat{a}, \hat{m}) . Assume that

- i) $(a, m) \mapsto \int_{\mathbb{R}^d} H(t, x, a, m, p) dx$ is convex for each $t \in [0, T]$,
- ii) $m \mapsto \int_{\mathbb{R}^d} g(T, x, m) dx$ is convex.

If, for all $w^j \in \mathcal{A}_d$ and for all $j \in \llbracket M \rrbracket$, we have that

$$\int_{\mathbb{R}^d} \int_0^T D_{aj} \hat{H}(t, x, p) \cdot w^j(t, x) dt dx = 0, \quad (4.22)$$

then \hat{a} solves (OC₀).

Before proving Theorem 4.2 we give an equivalent, but more useful, statement of the convexity assumptions *i*) and *ii*).

Proposition 4.2. *Conditions *i*) and *ii*) of Theorem 4.2 are satisfied if and only if $\bar{\Lambda}$ is positive semidefinite.*

Proof. Since g is linear in m , condition *ii*) holds. The quadratic dependence of H on a yields convexity of in the first argument of *i*). Convexity in the second argument holds if, for all densities m and m' (in $L^2(\mathbb{R}^d)$ and such that the measures they induce are in \mathcal{M}_2) and for all $\alpha \in (0, 1)$,

$$\begin{aligned} \int_{\mathbb{R}^d} H(t, x, a, \alpha m + (1 - \alpha)m', p) dt dx &\leq \alpha \int_{\mathbb{R}^d} H(t, x, a, m, p) dt dx \\ &\quad + (1 - \alpha) \int_{\mathbb{R}^d} H(t, x, a, m', p) dt dx, \end{aligned} \quad (4.23)$$

which is equivalent to

$$\begin{aligned} 0 &\geq \int_{\mathbb{R}^d} (\alpha m(t, x) + (1 - \alpha)m'(t, x))^T \bar{\Lambda} (\alpha m(t, x) + (1 - \alpha)m'(t, x)) \\ &\quad - \alpha m^T(t, x) \bar{\Lambda} m(t, x) - (1 - \alpha)m'^T(t, x) \bar{\Lambda} m'(t, x) dx \\ &= (\alpha^2 - \alpha) \int_{\mathbb{R}^d} (m(t, x) - m'(t, x))^T \Lambda (m(t, x) - m'(t, x)) dx. \end{aligned} \quad (4.24)$$

Since $\alpha < 1$, the inequality (4.24) holds for all m and m' if and only if $\bar{\Lambda}$ is positive semidefinite. \square

Remark 4.3. The weight matrix Λ in the example referred to in Remark 4.2 corresponds to

$$\bar{\Lambda} = \begin{bmatrix} 1 & \lambda/2 \\ \lambda/2 & 1 \end{bmatrix}, \quad (4.25)$$

which is positive semidefinite if and only if $|\lambda| \leq 2$.

Proof. (Proof of Theorem 4.2) The derivatives $D_m \hat{H}$ and $D_m \hat{g}$ are found by perturbing \hat{m} . Let m' be some other probability density on \mathbb{R}^d , then

$$m'(t, x) = \hat{m}(t, x) + \epsilon h'(t, x) + o(\|h'\|^2). \quad (4.26)$$

for some h' . We get

$$\begin{aligned}
H(t, x, \hat{a}, m', p) - H(t, x, \hat{a}, \hat{m}, p) &= \frac{1}{2} |\hat{a}(t, x)|^2 \cdot \epsilon h'(t, x) + \hat{m}^T(t, x) \bar{\Lambda} \epsilon h'(t, x) \\
&\quad + \epsilon h'^T(t, x) \bar{\Lambda} \hat{m}(t, x) + \epsilon^2 h'^T(t, x) \bar{\Lambda} h'(t, x) \\
&\quad + \sum_{j=1}^M b(t, x, \hat{a}^j(t, x)) h'_j(t, x) \cdot \nabla p_j(t, x), \\
g(x, m^\epsilon) - g(x, \hat{m}) &= \Psi(x) \cdot h'(T, x).
\end{aligned} \tag{4.27}$$

Proposition 4.2 says that the convexity condition on the integrated Hamiltonian forces $\bar{\Lambda}$ to be a positive semidefinite matrix, hence symmetric. Symmetry of $\bar{\Lambda}$ yields

$$D_m \hat{H}(t, x, p) = \frac{1}{2} |\hat{a}(t, x)|^2 + 2 \hat{m}^T(t, x) \bar{\Lambda} \tag{4.28}$$

$$\begin{aligned}
&\quad + (b(t, x, \hat{a}^1(t, x)) \cdot \nabla p_1(t, x), \dots, b(t, x, \hat{a}^M(t, x)) \cdot \nabla p_M(t, x)) \\
D_m \hat{g}(T, x) &= \Psi(x).
\end{aligned} \tag{4.29}$$

Let $\varphi^\epsilon(t, x, p) := \varphi(t, x, a_\epsilon, m^\epsilon, p)$ and $\hat{\varphi}(t, x, p) := \varphi(t, x, \hat{a}, \hat{m}, p)$ for $\varphi \in \{f, H, g\}$. Then

$$\begin{aligned}
f^\epsilon(t, x) - \hat{f}(t, x) &= H^\epsilon(t, x, p) - \hat{H}(t, x, p) - \sum_{j=1}^M b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) \cdot \nabla p_j(t, x) \\
&\quad + \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x)) - b(t, x, a_\epsilon^j(t, x))) m_j^\epsilon(t, x) \cdot \nabla p_j(t, x) \\
&= H^\epsilon(t, x, p) - \hat{H}(t, x, p) \\
&\quad - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x).
\end{aligned} \tag{4.30}$$

Assume that the control set A is convex and let, for some a and \hat{a} in \mathcal{A}_d^M ,

$$a_\epsilon(t, x) := \epsilon a(t, x) + (1 - \epsilon) \hat{a}(t, x), \quad \epsilon \in (0, 1). \tag{4.31}$$

Let the components of m^ϵ and \hat{m} satisfy the constraints of (OC₀) with a_ϵ and \hat{a} respectively. Then each component η_j of $\eta := m^\epsilon - \hat{m}$ satisfies the equation

$$\begin{cases} \frac{\partial \eta_j}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T \eta_j)(t, x)] - \nabla \cdot (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)), \\ \eta_j(0, x) = 0, \end{cases} \tag{4.32}$$

where $\kappa_\epsilon^j(t, x) = (b(t, x, a_\epsilon^j(t, x)) - b(t, x, \hat{a}^j(t, x))) m_j^\epsilon(t, x)$. The sufficiency is given by the

assumptions of the theorem and partial integration,

$$\begin{aligned}
& J_{0,\bar{\Lambda}}(a_\epsilon) - J_{0,\bar{\Lambda}}(\hat{a}) \\
&= \int_{\mathbb{R}^d} \int_0^T \left(f^\epsilon(t, x) - \hat{f}(t, x) \right) dt + g^\epsilon(x) - \hat{g}(x) dx \\
&\geq \int_{\mathbb{R}^d} \int_0^T \left(H^\epsilon(t, x, p) - \hat{H}(t, x, p) - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) \right) dt \\
&\quad + D_m \hat{g}(x) \cdot \eta(T, x) dx \\
&\geq \int_0^T \int_{\mathbb{R}^d} D_m \hat{H}(t, x, p) \cdot \eta(t, x) + \sum_{j=1}^M D_{a^j} \hat{H}(t, x, p) \cdot (a_\epsilon^j(t, x) - \hat{a}^j(t, x)) \\
&\quad - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) \\
&\quad + \frac{\partial p}{\partial t}(t, x) \cdot \eta(t, x) + p(t, x) \cdot \frac{\partial \eta}{\partial t}(t, x) dx dt \\
&= \int_0^T \int_{\mathbb{R}^d} \frac{1}{2} |\hat{a}(t, x)|^2 \cdot \eta(t, x) + 2\hat{m}^T(t, x) \bar{\Lambda} \eta(t, x) + \sum_{j=1}^M b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) \cdot \nabla p_j(t, x) \\
&\quad - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) + \frac{\partial p}{\partial t}(t, x) \cdot \eta(t, x) \\
&\quad + \sum_{j=1}^M p_j(t, x) \left(\frac{1}{2} \text{Tr} [\nabla^2 (\sigma^T \sigma \eta_j)(t, x)] - \nabla \cdot (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)) \right) dx dt \\
&= \int_0^T \int_{\mathbb{R}^d} \left(\frac{1}{2} |\hat{a}(t, x)|^2 + 2\hat{m}^T(t, x) \bar{\Lambda} + \frac{\partial p}{\partial t}(t, x) + \frac{1}{2} \text{Tr} [\sigma \sigma^T \nabla^2 p(t, x)] \right. \\
&\quad \left. + (b(t, x, \hat{a}^1(t, x)) \cdot \nabla p_1(t, x), \dots, b(t, x, \hat{a}^M(t, x)) \cdot \nabla p_M(t, x)) \right) \cdot \eta(t, x) dx dt
\end{aligned} \tag{4.33}$$

Since p solves (4.21), $J_{0,\bar{\Lambda}}(a_\epsilon) - J_{0,\bar{\Lambda}}(\hat{a}) \geq 0$ for all convex perturbations a_ϵ of \hat{a} . In the case of a control sets A which is not convex the proof can be carried out in similar fashion by replacing the convex perturbation a_ϵ by a spike variation. \square

Note that if

$$\hat{a}^j(t, x) = - (D_{a^j} b(t, x, \hat{a}^j(t, x))) \cdot \nabla p_j(t, x). \tag{4.34}$$

the optimality condition $\int_{\mathbb{R}^d} \int_0^T D_{a^j} \hat{H}(t, x) \cdot w^j(t, x) dt dx = 0$ is satisfied. In the case of linear dynamics, (4.34) is the well known solution $\hat{a}^j(t, x) = -\nabla p_j(t, x)$. The opposite direction of Proposition 4.1 can now be proven.

Proposition 4.3. *If \hat{a} solves dMFT-M with $r = 0$, \hat{m} satisfies the constraint in (OC₀) given \hat{a} and p satisfies the adjoint equation (4.21), then \hat{a} solves (OC₀).*

Proof. A proof is found in Appendix 6.3. \square

4.5. An optimal control problem equivalent to dMFT-M, $r > 0$

Let $r > 0$ and consider the following optimization problem.

$$\begin{aligned} & \text{minimize} \quad J_{r,\bar{\Lambda}}(a) \\ & \text{subject to} \quad \begin{cases} \frac{\partial m_j}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T m_j)(t, x)] - \nabla \cdot (b(t, x, a^j(t, x)) m_j(t, x)), \\ m_j(0, x) = \text{density of } \mu_0^j, \end{cases} \quad (OC_r) \\ & \quad a \in \mathcal{A}_d^M \end{aligned}$$

where

$$\begin{aligned} J_{r,\bar{\Lambda}}(a) := & \int_{\mathbb{R}^d} \left[\int_0^T \left(\frac{1}{2} |a(t, x)|^2 \cdot m(t, x) + G_{\phi_r}[m]^T(t, x) \bar{\Lambda} m(t, x) \right) dt \right. \\ & \left. + \Psi(x) \cdot m(T, x) \right] dx, \end{aligned} \quad (4.35)$$

and

$$G_{\phi_r}[m](t, x) := \left(\int_{\mathbb{R}^d} \phi_r(x - y) m_1(t, y) dy, \dots, \int_{\mathbb{R}^d} \phi_r(x - y) m_M(t, y) dy \right). \quad (4.36)$$

As in the previous section, we will find a solution to dMFT-M by solving an equivalent optimal control problem. An analog result to Proposition 4.1 holds.

Proposition 4.4. *If \hat{a} solves (OC_r) , then \hat{a} is a solution to dMFT-M when $r > 0$.*

Proof. The proof is found in Appendix 6.4. □

An optimal control to (OC_r) will be characterized through a sufficient maximum principle. Let

$$f(t, x, a, m) := \frac{1}{2} |a(t, x)|^2 \cdot m(t, x) + G_{\phi_r}[m](t, x)^T \bar{\Lambda} m(t, x) \quad (4.37)$$

and let g be as in the previous section.

Theorem 4.3. *Let $\hat{a} \in \mathcal{A}_d^M$, let*

$$H(t, x, a, m, p) := f(t, x, a, m) + \sum_{j=1}^M b(t, x, a^j(t, x)) m_j(t, x) \cdot \nabla p_j(t, x), \quad (4.38)$$

and let p solve the adjoint equation

$$\begin{cases} \frac{\partial p}{\partial t}(t, x) = - \left(\frac{1}{2} |\hat{a}(t, x)|^2 + 2 G_{\phi_r}[\hat{m}](t, x)^T \bar{\Lambda} \right. \\ \quad \left. + (b(t, x, \hat{a}^1(t, x)) \cdot \nabla p_1(t, x), \dots, b(t, x, \hat{a}^M(t, x)) \cdot \nabla p_M(t, x)) \right. \\ \quad \left. + \frac{1}{2} \text{Tr} [\sigma \sigma^T \nabla^2(p)(t, x)] \right), \\ p(T, x) = D_m \hat{g}(x), \end{cases} \quad (4.39)$$

where $D_m \hat{g}$ is the Gâteaux derivative of g with respect to m in the point \hat{m} . Assume that

- i) $(a, m) \mapsto \int_{\mathbb{R}^d} H(t, x, a, m, p) dx$ is convex for all $t \in [0, T]$,
- ii) $m \mapsto \int_{\mathbb{R}^d} g(T, x, m) dx$ is convex.

If, for all $w^j \in \mathcal{A}_d$ and for all $j \in \llbracket M \rrbracket$, it holds that

$$\int_{\mathbb{R}^d} \int_0^T D_{a^j} \hat{H}(t, x, p) \cdot w^j(t, x) dt dx = 0, \quad (4.40)$$

then \hat{a} solves (OC_r).

The essential difference between the setup of the theorem above and Theorem 4.2 is the Hamiltonian H . So before proving Theorem 4.3, let us calculate the m -derivative of H .

Lemma 4.1. *The Gâteaux derivative of H with respect to m at (\hat{a}, \hat{m}) applied to η is*

$$\begin{aligned} D_m \hat{H}[\eta](t, x, p) &= \frac{1}{2} |\hat{a}(t, x)|^2 \cdot \eta(t, x) + G_{\phi_r}[\hat{m}]^T(t, x) \bar{\Lambda} \eta(t, x) + G_{\phi_r}[\eta](t, x) \bar{\Lambda} \hat{m}(t, x) \\ &\quad + \sum_{j=1}^M b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) \cdot \nabla p_j(t, x). \end{aligned} \quad (4.41)$$

Proof. The same perturbation that was made in the proof of Proposition 4.3 (see Appendix 6.3) is used to get

$$\begin{aligned} H(t, x, \hat{a}, m', p) - H(t, x, \hat{a}, \hat{m}, p) &= \sum_{j=1}^N \left(\frac{1}{2} |\hat{a}^j(t, x)|^2 + b(t, x, \hat{a}^j(t, x)) \cdot \nabla p_j(t, x) \right) \epsilon h'_j(t, x) \\ &\quad + G_{\phi_r}[\hat{m}]^T(t, x) \bar{\Lambda} \epsilon h'(t, x) + G_{\phi_r}[\epsilon h']^T(t, x) \bar{\Lambda} \hat{m}(t, x) \\ &\quad + G_{\phi_r}[\epsilon h']^T(t, x) \bar{\Lambda} \epsilon h'(t, x). \end{aligned} \quad (4.42)$$

By (4.42) and the boundedness of the congestion weights we have, for all $(t, x) \in [0, T] \times \mathbb{R}^d$, that

$$\begin{aligned} \frac{|H(t, x, \hat{a}, m', p) - H(t, x, \hat{a}, \hat{m}, p) - D_m \hat{H}[\epsilon h'](t, x, p)|}{|\epsilon h'(t, x)|} &= \frac{|G_{\phi_r}[\epsilon h']^T(t, x) \bar{\Lambda} \epsilon h'(t, x)|}{|\epsilon h'(t, x)|} \\ &\leq C |G_{\phi_r}[\epsilon h'](t, x)|, \end{aligned} \quad (4.43)$$

for some constant $C > 0$. Since $G_{\phi_r}[\epsilon h'] \rightarrow 0$ as $\epsilon \rightarrow 0$ for all h' , the lemma follows. \square

Proof. (Proof of Theorem 4.3) Let a and \hat{a} be controls and let

$$a_\epsilon(t, x) := \epsilon a(t, x) + (1 - \epsilon) \hat{a}(t, x), \quad \epsilon \in (0, 1). \quad (4.44)$$

Let the components of m^ϵ and \hat{m} satisfy the constraints of (OC_r) with a_ϵ and \hat{m} respectively. Each component η_j of $\eta := m^\epsilon - \hat{m}$ then solves the equation

$$\begin{cases} \frac{\partial \eta_j}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\sigma^T \sigma(t, x) \nabla^2 \eta_j(t, x)] - \nabla \cdot (b(t, x, \hat{a}^j(t, x)) \eta_j(t, x) + \kappa_\epsilon^j(t, x)), \\ \eta_j(0, x) = 0. \end{cases} \quad (4.45)$$

Note that

$$f^\epsilon(t, x) - \hat{f}(t, x) = H^\epsilon(t, x, p) - \hat{H}(t, x, p) - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x))\eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) \quad (4.46)$$

and by symmetry of ϕ_n ,

$$\int_{\mathbb{R}^d} G_{\phi_r}[\hat{m}](t, x) \bar{\Lambda} \eta(t, x) dx = \int_{\mathbb{R}^d} G_{\phi_r}[\eta](t, x) \bar{\Lambda} \hat{m}(t, x) dx. \quad (4.47)$$

Sufficiency is given by the assumptions of the theorem and partial integration,

$$\begin{aligned} & J_{r, \bar{\Lambda}}(a_\epsilon) - J_{r, \bar{\Lambda}}(\hat{a}) \\ &= \int_{\mathbb{R}^d} \int_0^T f^\epsilon(t, x) - \hat{f}(t, x) dt + g^\epsilon(x) - \hat{g}(x) dx \\ &\geq \int_{\mathbb{R}^d} \int_0^T H^\epsilon(t, x, p) - \hat{H}(t, x, p) - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x))\eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) dt \\ &\quad + D_m \hat{g}_M(x) \cdot \eta(T, x) dx \\ &\geq \int_0^T \int_{\mathbb{R}^d} D_m \hat{H}[\eta](t, x, p) + \sum_{j=1}^M D_{a^j} \hat{H}(t, x, p) \cdot (a_\epsilon^j(t, x) - \hat{a}^j(t, x)) \\ &\quad - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x))\eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) \\ &\quad + \frac{\partial p}{\partial t}(t, x) \cdot \eta(t, x) + p(t, x) \cdot \frac{\partial \eta}{\partial t}(t, x) dx dt \\ &= \int_0^T \int_{\mathbb{R}^d} \sum_{j=1}^M \left(\frac{1}{2} |\hat{a}^j(t, x)|^2 + b(t, x, \hat{a}^j(t, x)) \cdot \nabla p_j(t, x) \right) \eta_j(t, x) \\ &\quad + G_{\phi_r}[\hat{m}]^T(t, x) \bar{\Lambda} \eta(t, x) + G_{\phi_r}[\eta]^T(t, x) \bar{\Lambda} \hat{m}(t, x) \\ &\quad - \sum_{j=1}^M (b(t, x, \hat{a}^j(t, x))\eta_j(t, x) + \kappa_\epsilon^j(t, x)) \cdot \nabla p_j(t, x) + \frac{\partial p}{\partial t}(t, x) \cdot \eta(t, x) \\ &\quad + \sum_{j=1}^M p_j(t, x) \left(\frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T \eta_j)(t, x)] - \nabla \cdot (b(t, x, \hat{a}^j(t, x))\eta_j(t, x) + \nabla \cdot \kappa_\epsilon^j(t, x)) \right) dx dt \\ &= \int_0^T \int_{\mathbb{R}^d} \left(\frac{1}{2} |\hat{a}(t, x)|^2 + 2G_{\phi_r}[\hat{m}]^T(t, x) \bar{\Lambda} + \frac{1}{2} \text{Tr} [\sigma \sigma^T \nabla^2 p(t, x)] \right. \\ &\quad \left. + (\hat{a}^1(t, x) \cdot \nabla p_1(t, x), \dots, \hat{a}^M(t, x) \cdot \nabla p_M(t, x)) + \frac{\partial p}{\partial t}(t, x) \right) \cdot \eta(t, x) dx dt \end{aligned} \quad (4.49)$$

By the adjoint equation (4.39) and the symmetry (4.47), $J_{r,\bar{\Lambda}}(a_\epsilon) - J_{r,\bar{\Lambda}}(\hat{a}) \geq 0$ for all convex perturbations a_ϵ of \hat{a} . In the case of a control sets A which is not convex the proof can be carried out in similar fashion by replacing the convex perturbation a_ϵ by a spike variation. \square

No property of $\bar{\Lambda}$ except boundedness in norm was used in the proof of the maximum principle. The following proposition identifies all matrices $\bar{\Lambda}$ that satisfy conditions *i*) and *ii*) of Theorem 4.3.

Proposition 4.5. *Conditions *i*) and *ii*) of Theorem 4.3 are satisfied if and only if*

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi_r(x-y)(m(t,y) - m'(t,y))^T \bar{\Lambda}(m(t,x) - m'(t,x)) dy dx \geq 0, \quad \forall t \in [0, T]. \quad (4.50)$$

for all m and m' (in $L^2(\mathbb{R}^d)$) such that the measures induced are in \mathcal{M}_2).

Proof. Since g is linear in m , condition *ii*) holds. The quadratic dependence of H on a yields convexity of in the first argument of *i*). Convexity in the second argument of *i*) holds if, for all densities m and m' (in $L^2(\mathbb{R}^d)$) and such that the induced measures are in \mathcal{M}_2), and for all $\alpha \in (0, 1)$, it holds that

$$\begin{aligned} \int_{\mathbb{R}^d} H(t, x, a, \alpha m + (1 - \alpha)m', p) dt dx &\leq \alpha \int_{\mathbb{R}^d} H(t, x, a, m, p) dt dx \\ &+ (1 - \alpha) \int_{\mathbb{R}^d} H(t, x, a, m', p) dt dx. \end{aligned} \quad (4.51)$$

The inequality above can be rearranged into

$$\begin{aligned} 0 &\geq (\alpha^2 - \alpha) \int_{\mathbb{R}^d} G_{\phi_r}[m - m'](t, x) \bar{\Lambda}(m(t, x) - m'(t, x)) dx \\ &= (\alpha^2 - \alpha) \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi_r(x-y)(m(t,y) - m'(t,y))^T \bar{\Lambda}(m(t,x) - m'(t,x)) dy dx \right). \end{aligned} \quad (4.52)$$

The fact that $(\alpha^2 - \alpha) < 0$ concludes the proof. \square

Note that as $r \rightarrow 0$, condition (4.50) reduces to the condition in Proposition 4.2. As in the case $r = 0$ the opposite direction of Proposition 4.3 can now be proven.

Proposition 4.6. *If \hat{a} solves dMFT-M when $r > 0$, \hat{m} satisfies the constraints of (OC_r) with strategy \hat{a} and p satisfies the adjoint equation (4.39), then \hat{a} solves (OC_r) .*

Proof. The proof is found in Appendix 6.5. \square

5. Numerical example

With the following numerical example we want to illustrate the difference between a crowd acting under local and one acting under nonlocal congestion. We consider the following simple

pedestrian model on the one-dimensional torus \mathbb{T} ,

$$\begin{aligned} & \underset{a \in \mathcal{A}_d}{\text{minimize}} && \int_{\mathbb{T}} \int_0^T \left\{ \frac{a^2(t, x)}{2} + C \int_{\mathbb{T}} \phi_r(x - y) m(t, y) dy \right\} m(t, x) dt + \Psi(x) m(T, x) dx \\ & \text{subject to} && \frac{\partial m}{\partial t}(t, x) = \frac{1}{2} \frac{\partial^2 m}{\partial x^2}(t, x) - \frac{\partial}{\partial x}(a(t, x) m(t, x)), \\ & && m(0, x) = m_0(x). \end{aligned} \tag{5.1}$$

To make the comparison we also consider a version of (5.1) where the pedestrians act under local congestion,

$$\begin{aligned} & \underset{a \in \mathcal{A}_d}{\text{minimize}} && \int_{\mathbb{T}} \int_0^T \left\{ \frac{a^2(t, x)}{2} + C m(t, x) \right\} m(t, x) dt + \Psi(x) m(T, x) dx \\ & \text{subject to} && \frac{\partial m}{\partial t}(t, x) = \frac{1}{2} \frac{\partial^2 m}{\partial x^2}(t, x) - \frac{\partial}{\partial x}(a(t, x) m(t, x)), \\ & && m(0, x) = m_0(x). \end{aligned} \tag{5.2}$$

The PDE-condition in (5.1) and (5.2) corresponds to the dynamics of a pedestrian that controls its velocity but is disturbed by white noise,

$$dX_t = a(t, X_t) dt + dW_t. \tag{5.3}$$

The constant C has been introduced to reweight the contribution of second term in the congestion risk. By up-weighting this term, we emphasize the impact of the congestion preference, local or nonlocal, and the difference between the two crowds will become more clear.

To solve (5.1) and (5.2), we implement the gradient decent method (GDM) of [12]. A recent paper that also addresses the numerical solution to systems of this kind is [1].

5.1. Simulations and discussions

We let $T = 1$, $C = 500$ and m_0 and ϕ_r are set to the functions presented in Figure 1. Most pedestrians are initially gathered around $x = 0$ and they have an incentive to end up around $x = 0.5$ at time $t = 1$.

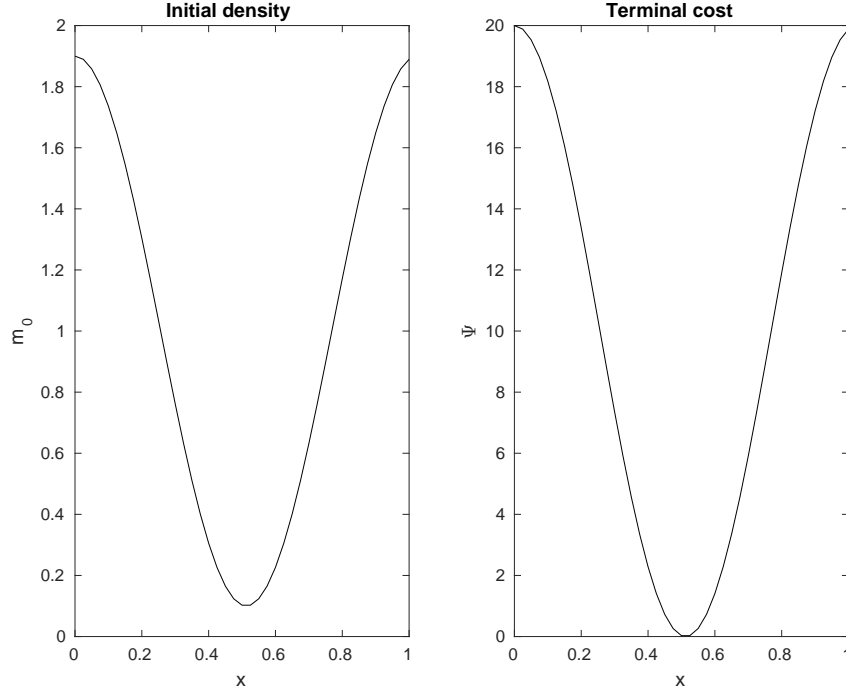


FIG 1. The initial density and terminal cost used in the simulations. Initially the pedestrians are crowded around $x = 0$ but they will quickly flatten the density to heed their congestion preferences. The low cost around $x = 0.5$ will give the pedestrians an incentive to end up around this part of the domain at $t = 1$.

The personal space of a pedestrian is modeled by the following density function on a neighborhood of radius $r = 0.2$,

$$\hat{\phi}_{0.2}(x) := 5\mathbb{I}_{[0, .2]}(x) \quad (5.4)$$

In the calculation $\hat{\phi}_{0.2}$ is smoothed with a mollifier, see (3.12). Note that

$$\int_{\mathbb{T}} \hat{\phi}_{0.2}(x - y)m(t, y)dy = 5\mathbb{P}(x - Y \in [0, 0.2]), \quad (5.5)$$

where Y is a random variable with density m . Thus the use of an indicator to model the personal space gives the following nice interpretation. The pedestrian acting under nonlocal congestion is affected by the probability of other pedestrians being closer than 0.2 from its own position. The averaging effect of a nonlocal congestion model is clear: the larger the personal space, the bigger neighborhood around the pedestrian is affecting it. If we shrink the neighborhood, the normalizing constant blows up.

The optimal controls for (5.1) and (5.2) are found by the GDM-scheme of [12]. The convergence of the congestion risk is presented in Figure 2.

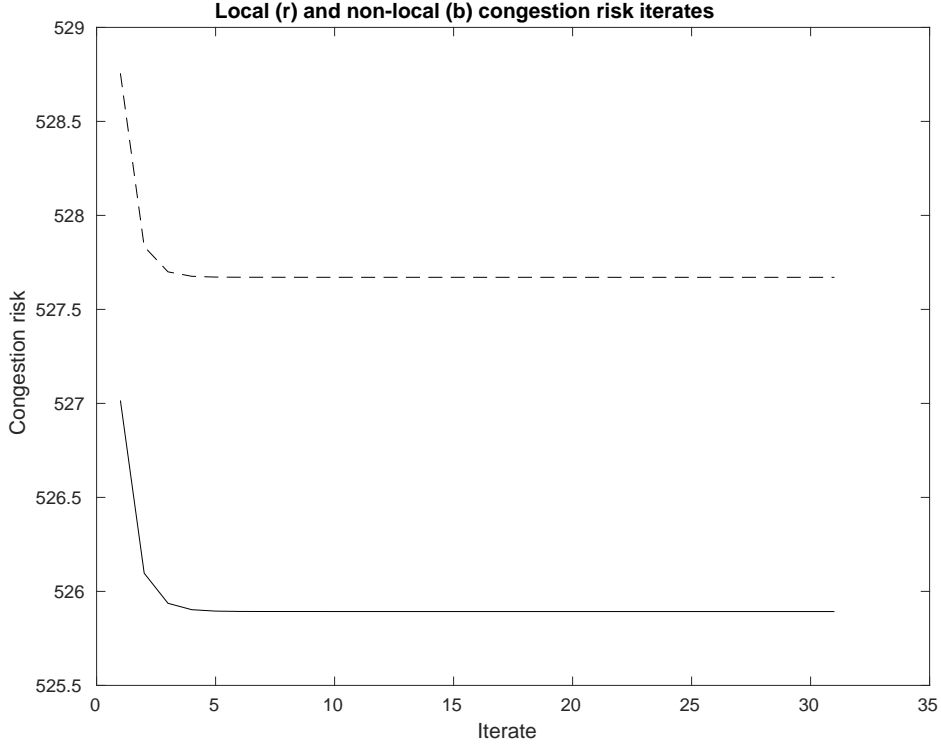


FIG 2. In each iteration of the GDM the control function a is updated. The method is run until the congestion risk, under local (dashed) and nonlocal (solid), has converged to a minimum.

In Figure 3, we display a comparison between the solutions to (5.1) and (5.2). The crowds behave similarly until the time begins to approach $t = 1$. The crowd acting under nonlocal congestion then gathers more densely in the low cost area. Since the congestion experienced by a pedestrian in the nonlocal model is an average over a larger neighborhood, it cares less about pointwise high densities and the benefits of reaching the low cost area around $x = 0.5$ affects the nonlocal model more strongly, resulting in a more concentrated density. This is visualized in Figure 4, where on the left the difference between congestion terms,

$$\underbrace{\int_{\mathbb{T}} \varphi_r(x-y) m_{\text{non-local}}(t, y) dy}_{\text{Nonlocal congestion term}} - \underbrace{m_{\text{local}}(t, x)}_{\text{Local ditto}}, \quad (5.6)$$

is plotted. On the right plot, we display

$$m_{\text{non-local}}(t, x) - m_{\text{local}}(t, x). \quad (5.7)$$

Note that even though the densities differ at $t = 1$, the two crowds generates approximately the same amount of congestion at that time!

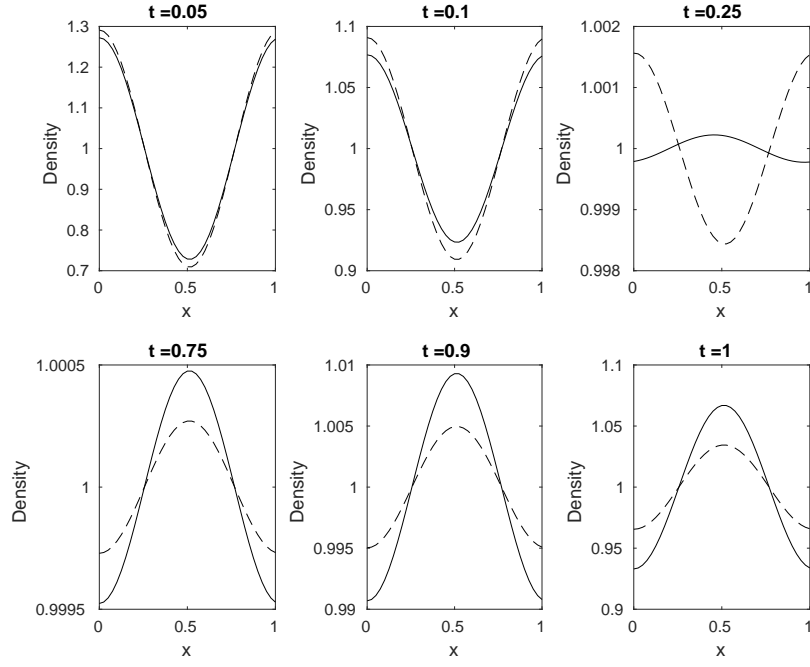


FIG 3. The optimally controlled density under local (dashed) and non-local (solid) congestion at six instants. Note the different scales on the y-axes!

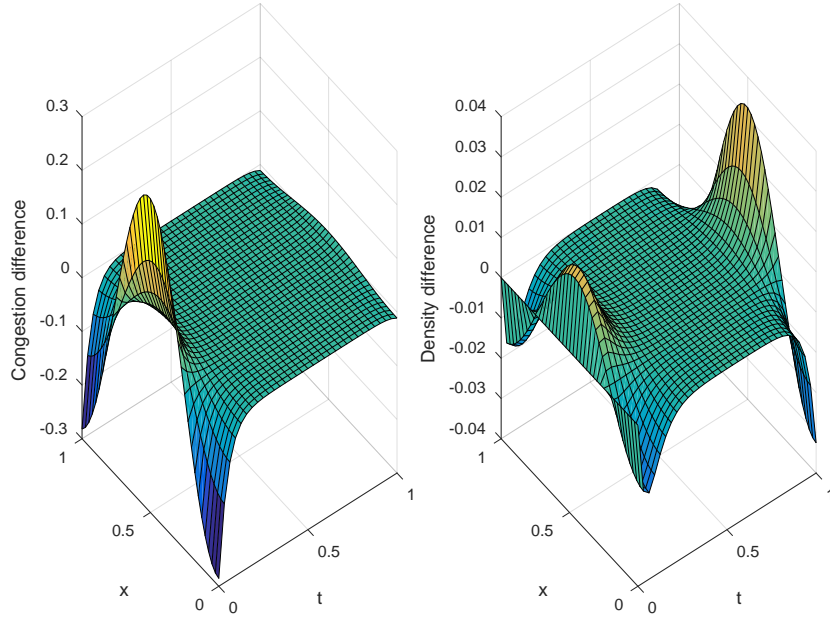


FIG 4. Differences (Non-local - Local) between the two crowds: congestion term (left plot) and density (right plot).

6. Appendix

6.1. Proof of Proposition 3.2

This proof is a straightfoward extention of results in [15] to systems of controlled stochastic differential equations.

Proof. Define for $f \in C_b^2(\mathbb{R}^d)$:

$$h_f : (\mathbb{R}^d)^N \rightarrow \mathbb{R} \quad (6.1)$$

$$x \mapsto \left\langle \frac{1}{N} \sum_{i=1}^N \delta_{x_i}, f \right\rangle = \frac{1}{N} \sum_{i=1}^N f(x_i). \quad (6.2)$$

Let $X_t = (X_t^{1,N}, \dots, X_t^{N,N}) \in (\mathbb{R}^d)^N$, by Itô's formula

$$dh_f(X_t) \quad (6.3)$$

$$= \frac{1}{N} \sum_{i=1}^N df(X_t^{i,N}) \quad (6.4)$$

$$= \frac{1}{N} \sum_{i=1}^N \left(\nabla f(X_t^{i,N}) \cdot dX_t^{i,N} + \frac{1}{2} \text{Tr} \left[\nabla^2 f(X_t^{i,N}) d\langle X^{i,N} \rangle_t \right] \right) \quad (6.5)$$

$$= \left(\frac{1}{2} \langle \text{Tr} [\sigma \sigma^T(t, \cdot) \nabla^2 f], \mu_t^N \rangle + \frac{1}{N} \sum_{i=1}^N \nabla f(X_t^{i,N}) \cdot b(t, X_t^{i,N}, a^N(t, X_t^{i,N})) \right) dt \quad (6.6)$$

$$+ \frac{1}{N} \sum_{i=1}^N \nabla f(X_t^{i,N}) \cdot \sigma(t, X_t^{i,N}) dW_t^{i,N}. \quad (6.7)$$

Thus

$$\begin{aligned} \mathbb{E}_N [h_f(X_t) \mid \mathcal{F}_s^N] &= h_f(X_s) + \mathbb{E}_N \left[\int_s^t \langle \nabla f(\cdot) \cdot b(u, \cdot, a^N(u, \cdot)), \mu_u^N \rangle du \mid \mathcal{F}_s^N \right] \\ &\quad + \mathbb{E}_N \left[\int_s^t \frac{1}{2} \langle \text{Tr} [\sigma \sigma^T(u, \cdot) \nabla^2 f], \mu_u^N \rangle du \mid \mathcal{F}_s^N \right]. \end{aligned} \quad (6.8)$$

Let $\varphi \in C^2(\mathbb{R}^d)$ be such that $\varphi(x) = |x|$ when $|x| \geq 1$, positive and satisfying $\|\nabla \varphi\|_\infty + \|\nabla^2 \varphi\|_\infty < \infty$ and let c_1, \dots, c_7 be positive constants. As a consequence of Proposition 3.1,

$$\mathbb{E}_N \left[\int_s^t \langle \nabla \varphi(\cdot) \cdot b(u, \cdot, a^N(u, \cdot)), \mu_u^N \rangle du \mid \mathcal{F}_s^N \right] \quad (6.9)$$

$$\leq \mathbb{E}_N \left[\int_s^t \langle c_1 \|\nabla \varphi\|_\infty |b(u, \cdot, a^N(u, \cdot))|, \mu_u^N \rangle du \mid \mathcal{F}_s^N \right] \quad (6.10)$$

$$\leq c_2(t - s), \quad (6.11)$$

and

$$\mathbb{E}_N \left[\int_s^t \frac{1}{2} \langle \sigma \sigma^T(u, \cdot) \nabla^2 \varphi, \mu_u^N \rangle du \mid \mathcal{F}_s^N \right] \quad (6.12)$$

$$\leq \mathbb{E}_N \left[\int_s^t \frac{1}{2} \langle c_3 \|\nabla^2 \varphi\|_\infty |\sigma \sigma^T(u, \cdot)|, \mu_u^N \rangle du \mid \mathcal{F}_s^N \right] \quad (6.13)$$

$$\leq c_4(t - s). \quad (6.14)$$

The inequalities yield that

$$t \mapsto h_\varphi(X_t) - c_5 t \quad (6.15)$$

is a supermartingale and by analogous reasoning,

$$t \mapsto h_\varphi(X_t) + c_5 t \quad (6.16)$$

is a submartingale. The proof can now be concluded in the same way as in [15]. We include the rest to complete the argument. For $\lambda > 1$, let B_λ be the closed ball of radius λ centered at the origin and let I be the indicator function on B_λ^C . Since φ dominates λI for all $x \in \mathbb{R}^d$, it holds for all $\delta > 0$

$$\mathbb{P}_N \left(\sup_{t \in [0, T]} h_I(X_t) > \delta \right) = \mathbb{P}_N \left(\sup_{t \in [0, T]} h_{\lambda I}(X_t) > \lambda \delta \right) \quad (6.17)$$

$$\leq \mathbb{P}_N \left(\sup_{t \in [0, T]} h_\varphi(X_t) > \lambda \delta \right) \quad (6.18)$$

$$\leq \frac{1}{\lambda \delta} \mathbb{E}_N [h_\varphi(X_t) + c_5 T] \quad (6.19)$$

$$\leq \frac{1}{\lambda \delta} \mathbb{E}_N [h_\varphi(X_0) + 2c_5 T]. \quad (6.20)$$

By (B1) and Hölder's inequality, it holds that

$$\mathbb{P}_N \left(\sup_{t \in [0, T]} h_{\lambda I}(X_t) > \delta \right) \leq \frac{c_6}{\lambda \delta}. \quad (6.21)$$

uniformly in N . Pick $\epsilon > 0$, μ_k and δ_k such that

$$\sum_{k=1}^{\infty} \mu_k = \epsilon, \quad \delta_k \rightarrow 0, \quad \delta_k > 0, \quad \mu_k > 0. \quad (6.22)$$

Let I_k be the indicator function on $B_{\lambda_k}^C$ where $\lambda_k = c_6/(\mu_k \delta_k)$. Then

$$\mathbb{P}_N \left(\bigcap_{k=1}^{\infty} \sup_{t \in [0, T]} h_{I_k}(X_t) \geq \delta_k \right) \leq \sum_{k=1}^{\infty} \frac{c_6}{\lambda_k \delta_k} = \epsilon. \quad (6.23)$$

Note that $Q = \bigcap_{k=1}^{\infty} Q_k$, where $Q_k = \{\mu \in \mathcal{P}_2(\mathbb{R}^d) \mid \langle \mu, I_k \rangle < \delta_k\}$ is a tight collection of measures since for all $\epsilon > 0$, there is a $\delta_k < \epsilon$ such that the compact set B_{λ_k} satisfies $\mu(\mathbb{R}^d \setminus B_{\lambda_k}) < \delta_k$.

Lemma

Q is closed.

Proof. Let $\{\mu_n\} \in Q_k$, $\mu \in \mathcal{P}_2(\mathbb{R}^d) \setminus Q_k$ and $\lim_{n \rightarrow \infty} \mu_n = \mu$. Since B_{λ_k} is compact there exists a smooth function $\alpha : \mathbb{R}^d \rightarrow \mathbb{R}$ which is constant 1 on B_{λ_k} and constant 0 on the complement to any open set containing B_{λ_k} . For construction of such a function, see [3]. In particular, α can be zero outside $B_{\lambda_k + \epsilon}^{\text{open}}$, the open ball of radius $\lambda_k + \epsilon$ and then

$$\langle \mu_n, (1 - \alpha)I_{B_{\lambda_k}^c} \rangle = \int_{B_{\lambda_k + \epsilon}^{\text{open}} \setminus B_{\lambda_k}} (1 - \alpha(x)) \mu_n(dx) + \int_{\mathbb{R}^d \setminus B_{\lambda_k + \epsilon}^{\text{open}}} (1 - \alpha(x)) \mu_n(dx) \quad (6.24)$$

$$\leq \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} \left((\lambda_k + \epsilon)^d - \lambda_k^d \right) + \delta_k, \quad (6.25)$$

while $\langle \mu, 1 - \alpha \rangle > \delta_k$. This contradicts the convergence and thus $\mu \in Q_k$ for all k . \square

The lemma gives compactness of Q , which implies compactness of $\mathcal{P}(Q)$ in $\mathcal{P}(\mathcal{P}(Q))$. Since $\{\mu_t^N, N \in \mathbb{N}\} \subset Q$ (\mathbb{P} -a.s.), Prokhorov's theorem gives tightness of $\{\mathcal{L}(\mu_t^N), N \in \mathbb{N}\}$. Using the definition of admissible controls, one can show that μ^N satisfies Kolmogorov's Continuity Theorem [15]. Tightness of $\{\mathcal{L}(\mu^N), N \in \mathbb{N}\}$ in $\mathcal{P}(\mathcal{M}_2)$ is then given by [6, Theorem 2, Chapter VI.4]. Skorokhod's Representation Theorem then yields convergence of μ^N to some limit (possibly random) measure μ in \mathcal{M}_2 . \square

6.2. Proof of Proposition 4.1

This proof extends the results in [12] to an arbitrary finite number of crowds.

Proof. Let the entries in Λ' be denoted by λ'_{jk} . For each $j \in \llbracket M \rrbracket$ and for some functions h_1, h_2 and h_3 the cost functional $J_{0, \Lambda'}(a)$ can be rewritten as

$$\begin{aligned} J_{0, \Lambda'}(a) &= \int_{\mathbb{R}^d} \int_0^T \sum_{j=1}^M \left(\frac{1}{2} |a^j(t, x)|^2 m_j(t, x) + m_j(t, x) \sum_{k=1}^M \lambda'_{jk} m_k(t, x) \right) dt \\ &\quad + \sum_{i=1}^M \Psi_j(x) m_j(T, x) dx \\ &= \int_{\mathbb{R}^d} \int_0^T \left(\frac{1}{2} |a^j(t, x)|^2 m_j(t, x) + \lambda'_{jj} m_j^2(t, x) + m_j(t, x) \sum_{\substack{k=1 \\ k \neq j}}^M 2\lambda'_{jk} m_k(t, x) \right. \\ &\quad \left. + h_1(a^{-j}, m_{-j}) \right) dt + \Psi_j(x) m_j(T, x) + h_2(m_{-j}) dx \\ &= J_{0, \Lambda}^{j, \text{det}}(a) + h_3(a^{-j}, m_{-j}). \end{aligned}$$

Since \hat{a} solves (OC₀), it holds that

$$J_{0, \Lambda'}(\hat{a}) \leq J_{0, \Lambda'}(a^j; \hat{a}^{-j}), \quad \forall a^j \in \mathcal{A}_d, \quad \forall j \in \llbracket M \rrbracket, \quad (6.26)$$

which, by (6.2), implies that

$$J_{0,\Lambda}^{j,\det}(\hat{a}) \leq J_{0,\Lambda}^{j,\det}(a^j; \hat{a}^{-j}), \quad \forall a^j \in \mathcal{A}_d, \quad \forall j \in \llbracket M \rrbracket. \quad (6.27)$$

Since (6.27) holds for all $j \in \llbracket M \rrbracket$, \hat{a} is a solution to dMFTG-M when $r = 0$. \square

6.3. Proof of Proposition 4.3

This proof is a variation of [11, Proposition 4.2.1] which extends it to an arbitrary finite number of crowds.

Proof. Let, for a given $\epsilon > 0$, a_ϵ^j be the first order perturbation of \hat{a}^j with some w^j such that

$$a_\epsilon^j(t, x) := \hat{a}^j(t, x) + \epsilon w^j(t, x) \in \mathcal{A}_d. \quad (6.28)$$

Let m_j^ϵ satisfy the constraints in (OC₀) with a_ϵ^j and let

$$m_j^\epsilon(t, x) := \hat{m}_j(t, x) + \epsilon h_j^\epsilon(t, x) + \mathcal{O}(h_j^{\epsilon^2}). \quad (6.29)$$

Then h_j^ϵ satisfies the equation

$$\begin{cases} \frac{\partial h_j^\epsilon}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T h_j^\epsilon)(t, x)] \\ \quad - \nabla \cdot \left(b(t, x, \hat{a}^j(t, x)) h_j^\epsilon(t, x) + \frac{b(t, x, a_\epsilon^j(t, x)) - b(t, x, \hat{a}^j(t, x))}{\epsilon} m_j^\epsilon(t, x) \right), \\ h_j^\epsilon(0, x) = 0. \end{cases} \quad (6.30)$$

Let $\mathcal{J}^j : \epsilon \rightarrow J_{0,\Lambda}^{j,\det}(a_\epsilon^j, \hat{a}^{-j})$. Since the functional is convex, \hat{a} solves dMFT-M when $r = 0$ if and only if

$$\frac{\partial \mathcal{J}^j}{\partial \epsilon}(0) = 0, \quad \forall w^j \text{ such that } \hat{a}^j + \epsilon w^j \in \mathcal{A}_d, \quad \forall j \in \llbracket M \rrbracket. \quad (6.31)$$

Condition (6.32) is equivalent to

$$\begin{aligned} \int_{\mathbb{R}^d} \left[\int_0^T \left(\hat{a}^j(t, x) \hat{m}_j(t, x) \cdot w^j(t, x) + \frac{1}{2} |\hat{a}^j(t, x)|^2 h_j^0(t, x) \right. \right. \\ \left. \left. + 2 \hat{m}_j(t, x) \lambda_{jj} h_j^0(t, x) + \sum_{k \neq j}^M \hat{m}_k(t, x) \lambda_{jk} h_k^0(t, x) \right) dt + \Psi_j(x) h_j^0(T, x) \right] dx = 0, \end{aligned} \quad (6.32)$$

where h_j^0 solves (6.30) in the limit $\epsilon \rightarrow 0$,

$$\begin{cases} \frac{\partial h_j^0}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T h_j^0)(t, x)] \\ \quad - \nabla \cdot \left(b(t, x, \hat{a}^j(t, x)) h_j^0(t, x) + D_{a^j} b(t, x, \hat{a}^j(t, x)) w^j(t, x) \hat{m}_j(t, x) \right), \\ h_j^0(0, x) = 0. \end{cases} \quad (6.33)$$

Since p satisfies the adjoint equation we have that $\Psi_j(x) = p_j(T, x)$ and

$$\begin{aligned}
& \int_{\mathbb{R}^d} p_j(T, x) h_j^0(T, x) dx \\
&= \int_{\mathbb{R}^d} \int_0^T \frac{\partial p_j}{\partial t}(t, x) h_j^0(t, x) + p_j(t, x) \frac{\partial h_j^0}{\partial t}(t, x) dt dx \\
&= \int_{\mathbb{R}^d} \int_0^T \left(-\frac{1}{2} |\hat{a}^j(t, x)|^2 \hat{m}_j(t, x) - 2 \hat{m}_j(t, x) \lambda_{jj} - \sum_{k \neq j}^M \lambda_{jk} \hat{m}_k(t, x) \right. \\
&\quad \left. - b(t, x, \hat{a}^j(t, x)) \cdot \nabla p_j(t, x) - \frac{1}{2} \text{Tr} [\sigma \sigma^T \nabla^2 p_j(t, x)] \right) h_j^0(t, x) \\
&\quad + \left(\frac{1}{2} \text{Tr} [\nabla^2 (\sigma \sigma^T h_j^0)(t, x)] - \nabla \cdot (b(t, x, \hat{a}^j(t, x)) h_j^0(t, x)) \right. \\
&\quad \left. - \nabla \cdot (D_{a^j} b(t, x, \hat{a}^j(t, x)) w^j(t, x) \hat{m}_j(t, x)) \right) p_j(t, x) dt dx.
\end{aligned} \tag{6.34}$$

Inserting (6.34) into (6.32) yields

$$\int_{\mathbb{R}^d} \int_0^T (\hat{a}^j(t, x) + D_{a^j} b(t, x, \hat{a}^j(t, x))^T \nabla p_j(t, x)) \hat{m}_j(t, x) \cdot w_j(t, x) dt dx = 0. \tag{6.35}$$

Note that since (6.35) holds for all $j \in \llbracket M \rrbracket$, \hat{a} satisfies the optimality condition in Theorem 4.2 and therefore \hat{a} is a solution to (OC₀) by Theorem 4.2. \square

6.4. Proof of Proposition 4.4

The proof extends the results in [12] to an arbitrary finite number of crowds and to nonlocal congestion.

Proof. Let the entries in Λ' be denoted by λ'_{jk} . For each $j \in \llbracket M \rrbracket$ and for some functions h_1, h_2 and h_3 , the cost functional $J_{r, \Lambda'}(a)$ can be rewritten as

$$\begin{aligned}
& J_{r, \Lambda'}(a) \\
&= \int_{\mathbb{R}^d} \left[\int_0^T \sum_{j=1}^M \left(\frac{1}{2} |a^j(t, x)|^2 m_j(t, x) \right. \right. \\
&\quad \left. \left. + \left(\int_{\mathbb{R}^d} \phi_r(x - y) m_j(t, y) dy \right) \sum_{k=1}^M \lambda'_{jk} m_k(t, x) \right) dt + \sum_{j=1}^M \Psi_j(x) m_j(T, x) \right] dx
\end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbb{R}^d} \left[\int_0^T \left(\frac{1}{2} |a^j(t, x)|^2 m_j(t, x) + \lambda'_{jj} \left(\int_{\mathbb{R}^d} \phi_r(x - y) m_j(t, y) dy \right) \lambda_{jj} m_j(t, x) \right. \right. \\
&\quad \left. \left. + \left(\int_{\mathbb{R}^d} \phi_r(x - y) m_j(t, y) dy \right) \sum_{\substack{k=1 \\ k \neq j}}^M 2\lambda'_{jk} m_k(t, x) \right. \right. \\
&\quad \left. \left. + h_1 \left(a^{-j}, m_{-j}, \int_{\mathbb{R}^d} \phi_r(x - y) m_{-j}(t, y) dy \right) \right) dt + \Psi_j(x) m_j(T, x) + h_2(m_{-j}) \right] dx \\
&= J_{r, \Lambda}^j(a^j, a^{-j}) + h_3 \left(a^{-j}, m_{-j}, \int_{\mathbb{R}^d} \phi_r(x - y) m_{-j}(t, y) dy \right).
\end{aligned} \tag{6.36}$$

Since \hat{a} solves (OC₀), it holds that

$$J_{n, \Lambda'}(\hat{a}) \leq J_{n, \Lambda'}(a^j, \hat{a}^{-j}), \quad \forall a^j \in \mathcal{A}_d, \quad \forall j \in \llbracket M \rrbracket, \tag{6.37}$$

which, by (6.36), implies that

$$J_{r, \Lambda}^{j, \det}(\hat{a}) \leq J_{r, \Lambda}^{j, \det}(a^j, \hat{a}^{-j}), \quad \forall a^j \in \mathcal{A}_d, \quad j \in \llbracket M \rrbracket. \tag{6.38}$$

Since (6.38) holds for all $j \in \llbracket M \rrbracket$, \hat{a} is a solution to dMFT-M when $r > 0$. \square

6.5. Proof of Proposition 4.6

This proof is a variation of [11, Proposition 4.2.1] which extends it to an arbitrary finite number of crowds and to nonlocal congestion.

Proof. Let, for a given $\epsilon > 0$, a_ϵ^j be the first order perturbation of \hat{a}^j for some arbitrary w^j such that

$$a_\epsilon^j(t, x) := \hat{a}^j(t, x) + \epsilon w^j(t, x) \in \mathcal{A}_d. \tag{6.39}$$

Let m_ϵ^j satisfy the constraints in (OC₀) with a_ϵ^j and let

$$m_\epsilon^j(t, x) := \hat{m}_j(t, x) + \epsilon h_\epsilon^j(t, x) + \mathcal{O}(h_\epsilon^{j2}). \tag{6.40}$$

Then h_ϵ^j satisfies the equation

$$\begin{cases} \frac{\partial h_\epsilon^j}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T h_\epsilon^j)(t, x)] \\ \quad - \nabla \cdot \left(b(t, x, \hat{a}^j(t, x)) h_\epsilon^j(t, x) + \frac{b(t, x, a_\epsilon^j(t, x)) - b(t, x, \hat{a}^j(t, x))}{\epsilon} m_\epsilon^j(t, x) \right), \\ h_\epsilon^j(0, x) = 0. \end{cases} \tag{6.41}$$

Let $\mathcal{J}^j : \epsilon \rightarrow J_{r, \Lambda}^{j, \det}(a_\epsilon^j, \hat{a}^{-j})$. Since the functional is convex, \hat{a} solves dMFT-M when $r > 0$ if and only if

$$\frac{\partial \mathcal{J}^j}{\partial \epsilon}(0) = 0, \quad \forall w^j \text{ such that } \hat{a}^j + \epsilon w^j \in \mathcal{A}_d, \quad \forall j \in \llbracket M \rrbracket. \tag{6.42}$$

Condition (6.42) is equivalent to

$$\begin{aligned} & \int_{\mathbb{R}^d} \left[\int_0^T \left(\hat{a}^j(t, x) \hat{m}_j(t, x) \cdot w^j(t, x) + \frac{1}{2} |\hat{a}^j(t, x)|^2 h_j^0(t, x) \right. \right. \\ & \quad \left. \left. + 2\lambda_{jj} \left(\int_{\mathbb{R}^d} \phi_r(x - y) \hat{m}_j(t, y) dy \right) h_j^0(t, x) \right. \right. \\ & \quad \left. \left. + \sum_{k \neq j}^M \lambda_{jk} \left(\int_{\mathbb{R}^d} \phi_r(x - y) \hat{m}_k(t, y) dy \right) h_k^0(t, x) \right) dt + \Psi_j(x) h_j^0(T, x) \right] dx = 0 \end{aligned} \quad (6.43)$$

where h_j^0 solves (6.41) in the limit $\epsilon \rightarrow 0$,

$$\begin{cases} \frac{\partial h_j^0}{\partial t}(t, x) = \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T h_j^0)(t, x)] \\ \quad - \nabla \cdot (b(t, x, \hat{a}^j(t, x)) h_j^0(t, x) + D_{a^j} b(t, x, \hat{a}^j(t, x)) w^j(t, x) \hat{m}_j(t, x)) , \\ h_j^0(0, x) = 0. \end{cases} \quad (6.44)$$

Recall that $\Lambda' = \frac{1}{2}(\Lambda + \text{diag}(\Lambda))$. Since p satisfies the adjoint equation, $\Psi_j(x) = p_j(T, x)$ and

$$\begin{aligned} & \int_{\mathbb{R}^d} p_j(T, x) h_j^0(T, x) dx \\ &= \int_{\mathbb{R}^d} \int_0^T \frac{\partial p_j}{\partial t}(t, x) h_j^0(t, x) + p_j(t, x) \frac{\partial h_j^0}{\partial t}(t, x) dt dx \\ &= \int_{\mathbb{R}^d} \int_0^T \left\{ -\frac{1}{2} |\hat{a}^j(t, x)|^2 - 2\lambda_{jj} \left(\int_{\mathbb{R}^d} \phi_r(x - y) \hat{m}_j(t, y) dy \right) \right. \\ & \quad \left. - \sum_{k \neq j}^M \lambda_{jk} \left(\int_{\mathbb{R}^d} \phi_r(x - y) \hat{m}_k(t, y) dy \right) - b(t, x, \hat{a}^j(t, x)) \cdot \nabla p_j(t, x) \right. \\ & \quad \left. - \frac{1}{2} \text{Tr} [\sigma \sigma^T \nabla^2 p_j(t, x)] \right\} h_j^0(t, x) + \left\{ \frac{1}{2} \text{Tr} [\nabla^2(\sigma \sigma^T h_j^0)(t, x)] \right. \\ & \quad \left. - \nabla \cdot (b(t, x, \hat{a}^j(t, x)) h_j^0(t, x) + D_{a^j} b(t, x, \hat{a}^j(t, x)) w^j(t, x) \hat{m}_j(t, x)) \right\} p_j(t, x) dt dx. \end{aligned} \quad (6.45)$$

Inserting (6.45) into (6.43) yields

$$\int_{\mathbb{R}^d} \int_0^T (\hat{a}^j(t, x) + D_{a^j} b(t, x, \hat{a}^j(t, x))^T \nabla p_j(t, x)) \cdot w_j(t, x) \hat{m}_j(t, x) dx dt = 0. \quad (6.46)$$

Note that this since (6.46) holds for all $j \in \llbracket M \rrbracket$, \hat{a} satisfies the optimality condition in Theorem 4.3 and therefore \hat{a} is a solution to (OC_r) by Theorem 4.3. \square

References

- [1] Achdou, Y. and Lauriere, M. [2016], ‘Mean Field Type Control with Congestion (II): An Augmented Lagrangian Method’, *arXiv preprint arXiv:1611.02023*.

- [2] Andersson, D. and Djehiche, B. [2011], ‘A maximum principle for sdes of mean-field type’, *Applied Mathematics & Optimization* **63**(3), 341–356.
- [3] Barden, D. and Thomas, C. B. [2003], *An introduction to differential manifolds*, World Scientific.
- [4] Bensoussan, A., Frehse, J., Yam, P. et al. [2013], *Mean field games and mean field type control theory*, Vol. 101, Springer.
- [5] Fischer, M. [2014], ‘On the connection between symmetric n -player games and mean field games’, *arXiv preprint arXiv:1405.1345*.
- [6] Gikhman, I. I. and Skorokhod, A. V. [2015], *The theory of stochastic processes I*, Springer.
- [7] Huang, M., Malhamé, R. P., Caines, P. E. et al. [2006], ‘Large population stochastic dynamic games: closed-loop mckean-vlasov systems and the nash certainty equivalence principle’, *Communications in Information & Systems* **6**(3), 221–252.
- [8] Kallenberg, O. [2006], *Foundations of modern probability*, Springer Science & Business Media.
- [9] Karatzas, I. and Shreve, S. [2012], *Brownian motion and stochastic calculus*, Vol. 113, Springer Science & Business Media.
- [10] Kolokoltsov, V. and Troeva, M. [2015], ‘On the mean field games with common noise and the mckean-vlasov spdes’, *arXiv preprint arXiv:1506.04594*.
- [11] Lachapelle, A. [2010], Quelques problèmes de transport et de contrôle en économie: aspects théoriques et numériques, PhD thesis, Université Paris Dauphine-Paris IX.
- [12] Lachapelle, A. and Wolfram, M.-T. [2011], ‘On a mean field game approach modeling congestion and aversion in pedestrian crowds’, *Transportation research part B: methodological* **45**(10), 1572–1589.
- [13] Lasry, J.-M. and Lions, P.-L. [2007], ‘Mean field games’, *Japanese Journal of Mathematics* **2**(1), 229–260.
- [14] Naldi, G., Pareschi, L. and Toscani, G. [2010], *Mathematical modeling of collective behavior in socio-economic and life sciences*, Springer Science & Business Media.
- [15] Oelschläger, K. [1985], ‘A law of large numbers for moderately interacting diffusion processes’, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **69**(2), 279–322.
- [16] Parthasarathy, K. R. [1967], *Probability measures on metric spaces*, Vol. 352, American Mathematical Soc.
- [17] Sznitman, A.-S. [1991], Topics in propagation of chaos, in ‘Ecole d’été de probabilités de Saint-Flour XIX1989’, Springer, pp. 165–251.
- [18] Tcheukam, A., Djehiche, B. and Tembine, H. [2016], Evacuation of multi-level building: Design, control and strategic flow, in ‘Control Conference (CCC), 2016 35th Chinese’, IEEE, pp. 9218–9223.
- [19] Yamada, T., Watanabe, S. et al. [1971], ‘On the uniqueness of solutions of stochastic differential equations’, *Journal of Mathematics of Kyoto University* **11**(1), 155–167.
- [20] Yong, J. and Zhou, X. Y. [1999], *Stochastic controls: Hamiltonian systems and HJB equations*, Vol. 43, Springer Science & Business Media.